

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 9/00, C07H 19/00, 21/00, C07K 1/00, 14/00, 17/00	A1	(11) International Publication Number: WO 97/20038 (43) International Publication Date: 5 June 1997 (05.06.97)
(21) International Application Number: PCT/US96/19277 (22) International Filing Date: 27 November 1996 (27.11.96) (30) Priority Data: 60/008,263 1 December 1995 (01.12.95) US (60) Parent Application or Grant (63) Related by Continuation US Not furnished (CIP) Filed on Not furnished (71) Applicant (for all designated States except US): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US). (71)(72) Applicant and Inventor: CHOW, Samson, A. [US/US]; 4133 Magna Carta Row, Calabasas, CA 91302 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): GOULAOUIC, Hélène [FR/FR]; 38, rue Paine, F-75012 Paris (FR).		(74) Agent: MAYFIELD, Denise, L.; Akin, Gump, Strauss, Hauer & Feld, L.L.P., Suite 1900, 816 Congress Avenue, Austin, TX 78701 (US). (81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: COMPOSITIONS AND METHODS FOR SITE-DIRECTED INTEGRATION INTO DNA (57) Abstract The present invention provides fusion proteins capable of integrating a donor DNA molecule into a target DNA molecule at or near a target nucleotide sequence. The fusion proteins comprise a retroviral integrase catalytic domain COOH-terminally coupled to a DNA binding protein domain having binding specificity for the target nucleotide sequence. Nucleic acids encoding same; vectors, expression systems, and host cells carrying nucleic acids encoding said fusion proteins; and methods of integrating a donor DNA molecule at or near a specific site on a target DNA molecule are provided. The integrating may result in a gene encoding a therapeutic to be introduced via gene therapy, or may result in an oncogene being inactivated, for example.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

COMPOSITIONS AND METHODS FOR SITE-DIRECTED INTEGRATION INTO DNA

5 The government owns certain rights in the present invention pursuant to grants from the Department of Energy (DE-FC03-87-ER60615) and the National Institutes of Health (R01 CA68859). The application claims priority to United States Patent Application Serial No. 60/008,263, filed December 1, 1995.

FIELD OF THE INVENTION

10

The present invention relates generally to molecular biological techniques for manipulating nucleic acid molecules. In particular, the present invention provides a fusion protein comprising an N-terminal integrase catalytic domain and a C-terminal nucleic acid binding domain having binding specificity for a target nucleic acid. The fusion protein is useful for site-specific integration of a donor nucleic acid into a target nucleic acid at or near the site of binding of the nucleic acid binding protein. Nucleic acids encoding the fusion protein, expression vectors, hosts, and methods of integrating a donor nucleic acid into a target nucleic acid are provided.

20

BACKGROUND OF THE INVENTION

Retroviral RNA is copied by the enzyme reverse transcriptase into a double-stranded linear viral DNA which is integrated into the host genome as a provirus. Integration of retroviral DNA into the host cell genome is an essential step during the life cycle of retroviruses (Varmus and Brown, 1989). Three factors are required for the integration process: the viral protein integrase, sequences at each end of the linear viral DNA, and a divalent metal ion cofactor. The human immunodeficiency virus type 1 integrase is encoded as a 32-kDa protein at the C-terminus of the Gag-Pol polyprotein which is processed into its individual components by the viral protease during budding. Integrase can be considered as having three domains, an N-terminal zinc finger domain, a central catalytic domain, and a C-terminal DNA binding domain.

30

The viral DNA precursor for the integration reaction is a linear double-stranded molecule. Two bases from each 3' end of the linear viral DNA are removed by

integrase such that the viral 3' ends are recessed by two bases from the 5' ends and terminate with the dinucleotide CA. A staggered cut is then made in the target DNA and the resulting overhanging 5'-P ends are covalently joined to the recessed 3'-OH ends of the viral DNA. For reviews of this concerted cleavage-joining reaction, see Brown
5 (1990), Goff (1992), and Vink and Plasterk (1993). This cleavage-ligation reaction produces a gapped intermediate; integration is completed by a gap repair process that remains to be characterized. In addition, integrase can carry out an *in vitro* reversal of the integration reaction, named disintegration, in which a branched DNA structure resembling an integration product is converted into two molecules resembling the initial
10 viral and target DNAs.

In vivo and *in vitro* studies show that integration of retroviral DNA can occur into many sites on target DNA (Craigie, 1992, and references therein). The process, however, is not entirely random; the frequency of use of specific sites varies
15 considerably, with some sites being preferred up to hundred times greater than random (Rohdewohld *et al.*, 1987; Vijaya *et al.*, 1986; Withers-Ward *et al.*, 1994). The mechanism that determines target site specificity is not well understood, and several factors have thus far been identified that can affect target site selection, including DNA and chromatin structure, DNA methylation, DNA sequences, and DNA-binding
20 proteins. Integration occurs preferentially into regions near DNase I-hypersensitive sites and transcriptionally active genes (Rohdewohld *et al.*, 1987; Vijaya *et al.*, 1986), and into runs of CpG islands modified by 5-methylation of cytosine (Kitamura *et al.*, 1992).

25 One factor important for target site selection that has been well characterized is chromatin structure. Nucleosomal DNA in the chromatin is preferred to nucleosome-free DNA, and integration tends to cluster in the exposed face of the major groove within the nucleosome core (Pruss *et al.*, 1994; Pryciak and Varmus, 1992). The basis for preferred integration in nucleosomes may be related to DNA distortion,
30 as DNA bending itself creates favored sites for integration (Muller and Varmus, 1994; Pruss *et al.*, 1994). Although sequence analysis of integration sites has only revealed

weak consensus sequences (Fitzgerald and Grandgenett, 1994; Grandgenett *et al.*, 1993), comparisons of the integration patterns in a DNA sequence *in vivo* and as a naked DNA *in vitro* show that the DNA sequence is also an important determinant in target site selection (Pryciak *et al.*, 1992; Pryciak and Varmus, 1992).

5

Another factor in target site selection is sequence- or structure-specific DNA binding proteins. Certain DNA-binding proteins, such as the yeast transcriptional repressor $\alpha 2$ and the *lac* repressor of *E. coli*, can prevent integration, presumably by steric hindrance (Muller and Varmus, 1994; Pryciak and Varmus, 1992). Unlike
10 histones and other proteins that stimulate integration by inducing DNA bends, certain DNA-binding proteins may promote integration by interacting with the integration machinery. The significance of such an interaction is illustrated by the position-specific integration of the yeast retrovirus-like element Ty3 (Sandmeyer *et al.*, 1990).

15

Integrase itself is a major factor in determining target site specificity. Integration reactions carried out with purified integrase or integration complexes isolated from virus-infected cells show similar patterns of target specificity. The C-terminal third of integrase, the least conserved region among retroviral integrases (Johnson *et al.*, 1986), possesses DNA-binding activity (Engelman *et al.*, 1994; Schauer
20 and Billich, 1992; Vink *et al.*, 1993; Woerner *et al.*, 1992). The DNA binding by the C-terminus does not show any sequence specificity, which led to its proposed role as the domain for binding target DNA, and this binding may partly explain the ability of integrase to insert viral DNA at sites with weak consensus sequences.

25

Directed integration has been reported by tethering integrase to a target DNA site, accomplished by use of a hybrid protein composed of the DNA-binding domain of λ repressor at the N-terminus and a full-length HIV-1 integrase at the C-terminus of the hybrid protein (Bushman, 1994). The hybrid protein mediates integration preferentially to target DNA containing λ operators. The integration sites are near the
30 λ operator on the same face of the DNA helix, indicating that the hybrid protein binds

to the operator and captures targets probably by looping out the intervening DNA (Bushman, 1994).

Various methods are currently being used in genetic engineering to enable the transfer and expression of genes into the genomes of cells and organisms. Genes have been transferred by incubating cells with DNA, possibly in the presence of chemicals such as polyions or calcium phosphate. Genetic material can also be injected into the nucleus or cytoplasm of cells or zygotes. Other methods include electroporation, liposome mediated gene insertion, asialoglycoprotein gene insertion, particle acceleration and viral transduction. The use of viruses in the transduction method has been shown to be very efficient when retroviruses are used. Foreign genes are inserted into either a replication defective or replication competent viral vector construct (usually as a plasmid), and are transferred into cells containing all the genes necessary for packaging and replication of the virus. Special cell lines ("helper" or viral packaging cells) have been constructed which enable defective (non-replication competent) viral vectors to be packaged into infectious particles or virions. The vectors themselves do not harbor the necessary genes for replication so that when the vectors infect cells, the vectors replicate using the enzymes in the viral particle to insert themselves into the host genome (chromosomes). The vectors should be unable to replicate further because the essential viral genes were left behind in the "helper" cell. This technique has been adopted and approved for the first human gene therapy trials, despite ongoing debate about the safety of such usages.

Retroviruses are now widely used as vectors for genetic engineering in higher eukaryotes and are considered to be promising vectors for gene therapy, owing to their natural aptitude for introducing foreign genes into cellular chromosomes (Mulligan, 1993). However, several features of current retroviral vectors limit their usefulness in gene therapy, including the limited size of their genome, their inability to infect nondividing cells, and their inability to target integration to a specific site (Mulligan, 1993; Shiramizu *et al.*, 1994; Temin, 1990). Indeed, the major shortcoming of retroviral vectors is their inability to target the DNA integration to a specific site. With

random integration, there is a risk of activating a proto-oncogene or inactivating a tumor suppressor gene in the target DNA.

5 There is a need in the art of molecular biology techniques for a method to integrate nucleic acids at a specific sequence. Because of the above problems, known procedures are not completely satisfactory, and persons skilled in the art have searched for improvements. The present inventors have carried out studies on target site selection to overcome these problems.

10 SUMMARY OF THE INVENTION

The present invention seeks to overcome these and other drawbacks inherent in the prior art by providing a fusion peptide having an N-terminal retroviral integrase catalytic domain covalently bonded to a C-terminal DNA binding moiety. Integration into a specific site is facilitated by the fusion protein since the DNA binding moiety provides the binding specificity for a particular site on a target DNA molecule and the integrase catalytic domain provides the catalytic machinery for accomplishing the integration. An aspect of the invention, therefore, is a fusion protein comprising a retroviral integrase catalytic domain COOH-terminally coupled to a DNA binding protein domain having binding specificity for a target nucleotide sequence, the fusion protein capable of integrating a donor DNA molecule into a target DNA molecule at or near the target nucleotide sequence.

"Integrase catalytic domain" is meant to include the sequence of amino acids from the catalytic domain of a retroviral integrase capable of carrying out disintegration, an *in vitro* reversal of the normal DNA strand transfer reaction. Generally speaking, the catalytic domain includes amino acids from about position 50 to about position 212, or about position 234, of the HIV-1 integrase (Cannon *et al.*, 1994). The catalytic domain is relatively conserved among retroviral integrases, and this region may be considered as applying to other retroviral integrases as well as HIV-1 integrase (Engelman and Craigie, 1992).

Disintegration is the reverse reaction of integration. In this reaction, a branched oligonucleotide substrate, or Y-mer, is resolved into its constituent donor and target double-stranded DNA components (see FIGS. 1-3 and brief description thereof). The disintegration substrate has the advantage that the site of integration into target DNA is predetermined and can be manipulated. The disintegration substrate is therefore particularly well suited for studies that benefit from a defined site of integration, such as investigations of protein-target DNA interactions during retroviral DNA integration.

The nucleotide sequence and structural requirements for disintegration are less stringent than those for 3' processing and strand transfer (Chow *et al.*, 1992). This characteristic allows genetic variants of integrase that lack detectable activity in 3' processing and strand transfer to retain disintegration activity (Bushman *et al.*, 1993; Engelman and Craigie, 1992; Leavitt *et al.*, 1993; van Gent *et al.*, 1992; Vincent *et al.*, 1993; Vink *et al.*, 1993). Thus, the disintegration assay has played an important role in locating the catalytic domain of integrase and is useful in mapping other functional domains of the protein (Chow and Brown, 1994).

A retroviral integrase may be human immunodeficiency virus type 1 or type 2, simian immunodeficiency virus, equine infectious anemia virus, feline immunodeficiency virus, caprine arthritis-encephalitis virus, bovine immunodeficiency virus, Mason-Pfizer monkey virus, mouse mammary tumor virus, intracisternal A particle, Rous sarcoma virus, bovine leukemia virus, human T-cell leukemia virus type I or II, reticuloendotheliosis virus, feline leukemia virus, murine leukemia virus or human spumaretrovirus, for example (see Engelman and Craigie, (1992), which reference is incorporated by reference herein in its entirety for this purpose, and references therein for amino acid sequences of integrase from these sources and for source information). A retroviral integrase may also be from avian myeloblastosis virus (Grandgenett *et al.*, 1993) or from visna virus (Katzman and Sudol, 1994). Retrotransposons, some eukaryotic and prokaryotic transposons, and the integrase of murine leukemia virus also share mechanistic features of HIV integration. Preferably,

the retroviral integrase catalytic domain is integrase from human immunodeficiency virus type 1 or type 2, or from feline immunodeficiency virus integrase.

A "DNA binding protein domain" or moiety is a functional amino acid sequence that has binding affinity and specificity for a particular nucleotide sequence in DNA. A DNA binding protein domain may include binding domains from: Cro repressor from phage lambda, cI repressor from phage lambda, Cro from phage 434, cI repressor from phage 434, P22 repressor, *E. coli* tryptophan repressor, *E. coli* CAP, P22 Arc, P22 Mnt, *E. coli* lactose repressor, tetracycline repressor from *E. coli*, MAT-a1-alpha2 from yeast, GAL4 from yeast, Polyoma Large T antigen, SV40 Large T antigen, adenovirus E1A, TFIIA from *Xenopus laevis*, or zinc finger DNA binding proteins. An example of a DNA binding protein domain is one having binding specificity for a target nucleotide sequence is LexA binding protein domain. A preferred target nucleotide sequence is the LexA consensus sequence, CTGTNNNNNNNNACAG, (SEQ ID NO:20) and a more preferred target nucleotide sequence is the LexA sequence, CTGTATGAGCATACAG, (SEQ ID NO:21).

The N-terminal integrase catalytic domain is covalently bonded at its carboxy terminus to a DNA binding protein domain, so that the DNA binding protein domain is at the carboxy terminus of the resultant fusion protein. The covalent bonding may be accomplished chemically by fusing the C-terminal carboxyl group of the integrase domain to the N-terminal amide group of the DNA binding moiety to form a peptide bond, but the fusion protein is more easily made by genetic engineering means, for example, by ligating nucleotide sequences together that encode the different moieties. One of skill in this art in light of the present disclosure would realize that some flexibility exists in the junction of the two protein domains, for example, a number of amino acids may be added or deleted as a consequence of cloning. However, it is important that the DNA binding domain nucleotide sequence be in the same reading frame as the nucleotide sequence encoding the integrase domain.

The fusion proteins of the present invention are useful for their capability of integrating a donor DNA molecule into a target DNA molecule at or near a target nucleotide sequence. This utility is very broad and includes the integration of genes encoding therapeutic products, or the integration of a piece of DNA for purposes of
5 disrupting a particular function, disrupting oncogene function, for example. By way of example, a preferred fusion protein has an amino acid sequence essentially as set forth in SEQ ID NO:23, or SEQ ID NO:25, SEQ ID NO:29, or SEQ ID NO:31, a combination thereof, or a biologically functional fragment thereof..

10 "Capable of integrating a donor DNA molecule into a target DNA molecule at or near the target nucleotide sequence" means that the donor DNA molecule may be integrated within a distance of about 30-50 base pairs or so from the target nucleotide sequence. The DNA binding domain, when bound to the nucleotide sequence for which it has affinity, will occupy about 30 nucleotides and therefore, the actual binding site
15 is unavailable for integration. Integration will preferably occur within about 30-50 base pairs of the DNA binding site, a distance affected in part by topology and flexibility of the fusion protein and the target DNA molecule.

The conditions for integration include temperatures for enzymatic activity to
20 occur, preferably at room or body temperature, keeping in mind that the reaction will occur more slowly at lower temperatures. A divalent metal cation is important for catalysis, preferably the cation is Mn(II) or Mg(II).

A fusion protein having an N-terminal integrase catalytic domain and a nucleic
25 acid binding domain at the C-terminus has several advantages over a construction where the nucleic acid binding domain is at the N-terminus of the fusion protein. For example, when the DNA encoding the fusion protein is introduced into the viral genome, placement of the DNA-binding protein at the N-terminus of integrase may affect the ability of viral protease to process the precursor polypeptide, leading to
30 defective viruses and nonfunctional proteins. It is therefore, an advantage to place the DNA-binding protein at the C-terminus of integrase.

When compared with the retroviral vectors currently available, the invention provides major improvements as a result of site-specific integration; i) safety - insertion of exogenous DNA will be directed towards innocuous regions of chromosomes, and away from essential genes, cancer-causing genes, or tumor suppressor genes, and ii) improved expression- insertion of exogenous DNA will be directed towards regions that are known for efficient and stable expression of genes.

"Donor DNA" is a linear double-stranded oligonucleotide with end sequences of about 15-35 nucleotides derived from the U5 or U3 ends of the retroviral long terminal repeat (LTR) (Varmus and Brown, 1989). The LTR contains regulatory sequences, such as promoter and enhancer sequences for gene expression, transcription initiation, and polyadenylation. Since the LTR sequence varies among different retroviruses, the exact sequence of the ends of the donor DNA will depend on the particular integrase used in the fusion construct. For instance, if the fusion protein comprises HIV-1 integrase and LexA protein, the sequences of the ends of the donor DNA will be constructed so as to mimic either the U5 or U3 end of the HIV-1 LTR. Although there is no consensus DNA sequence for the retroviral LTR, one invariant feature is a CA dinucleotide at positions 3 and 4 from the 3' end of the processed DNA strand. The donor DNA can be blunt-ended with the CA dinucleotide located 2 nucleotides from the 3' end of the processed strand. The donor DNA can also have a 5' extension, with the 3' end terminating with the CA dinucleotide.

The donor DNA may be a DNA molecule up to 10 kbp in length. In such a case, the donor DNA may contain the entire LTR (350 -700 bp) at both ends of the donor DNA. The sequence of the LTR corresponds to that of the retrovirus from which the integrase component of the fusion protein is obtained. Between the two LTRs, the donor DNA contains a psi sequence which is important for RNA packaging, and may contain a gene for therapeutic purposes (e.g. cystic fibrosis gene), or a reporter gene for selection (e.g. neomycin resistant gene) or for gene disruption, or a toxic gene for cell killing (e.g. ricin gene).

"Target DNA" is DNA that has a site recognizable by a DNA binding protein domain. A DNA molecule can be made into a target DNA by incorporation of nucleotides, the sequence of which is recognizable by a DNA binding protein domain. Incorporation of a sequence of nucleotides is most easily accomplished by restriction
5 enzyme digestion of a DNA, and ligation to a double stranded oligonucleotide having the particular sequence of nucleotides and having end linkers corresponding to the restriction enzyme used. Therefore, the target DNA is very broad, and includes any sequence where one would desire to incorporate a donor DNA molecule.

10 In certain aspects, the invention relates to a purified nucleic acid molecule consisting essentially of a nucleotide sequence encoding an integrase-DNA binding protein domain fusion protein, the protein having an amino acid sequence essentially as set forth in SEQ ID NOS:23, 25, 29 or 31. "Purified" nucleic acid molecule having a nucleotide sequence encoding an integrase-DNA binding protein domain fusion
15 protein, as used herein, means a fusion protein encoding nucleic acid molecule substantially free of nucleic acid molecules not encoding a fusion protein essentially as set forth in SEQ ID NOS:23, 25, 29 or 31. Preferably, the purified nucleic acid molecule is a DNA molecule wherein the nucleotide sequence is essentially as set forth in SEQ ID NOS:22, 24, 28, or 30.

20

The term "amino acid sequence essentially as set forth in SEQ ID NOS:23, 25, 29 or 31" means that the sequence substantially corresponds to a portion of SEQ ID NOS:23, 25, 29 or 31, and has relatively few amino acids which are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NOS:23, 25, 29 or
25 31. The term "biologically functional equivalent" is well understood in the art and is further defined as a protein having a sequence essentially as set forth in SEQ ID NOS:23, 25, 29 or 31, capable of integrating a donor DNA molecule into a target DNA molecule at or near a site specific to the DNA binding protein domain portion of the fusion protein. Accordingly, sequences which have between about 70% and about 80%;
30 or more preferably, between about 81% and about 90%; or even more preferably, between about 91% and about 99%; of amino acids that are identical or functionally

equivalent to the amino acids of SEQ ID NOS:23, 25, 29 or 31 will be sequences which are "essentially as set forth in SEQ ID NOS:23, 25, 29 or 31".

5 A further embodiment of the present invention is where the nucleic acid molecule has a nucleotide sequence as set forth in SEQ ID NOS:22, 24, 28, 30, a combination or a biologically functional fragment thereof. In some embodiments, the nucleic acid molecule is further defined as including a detectable label.

10 An embodiment of the present invention is a purified nucleic acid molecule that encodes an integrase-DNA binding moiety fusion protein. The fusion protein includes at a minimum an integrase catalytic domain covalently bonded to a DNA binding moiety and may have an amino acid sequence in accordance with SEQ ID NOS: 23, 25, 29, 31, a combination or a biologically functional fragment thereof. As used herein, the terms "nucleic acid molecule" may refer to a DNA or RNA molecule which has been
15 isolated free of total genomic DNA, or free of total RNA, of a particular species. Therefore, a "purified" nucleic acid molecule as used herein, refers to a nucleic acid molecule that contains an integrase catalytic domain-DNA binding moiety coding sequence, yet is isolated away from, or purified free from, total genomic DNA or total RNA, for example, total human genomic DNA. Included within the term "DNA
20 molecule", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like. The term "biologically functional" as used in the description of the present invention is defined as a capable of providing the site-directed integration of a nucleic acid into DNA as described in the present disclosure.

25

Another embodiment of the present invention is a purified nucleic acid molecule, further defined as including a nucleotide sequence in accordance with SEQ ID NOS:22, 24, 28 or 30. In a more preferred embodiment the purified nucleic acid segment consists essentially of the nucleotide sequence of SEQ ID NOS:22, 24, 28, 30,
30 or a combination thereof. Such nucleotide sequences are more particularly defined as being substantially free of nucleic acids not encoding the corresponding fusion protein.

Similarly, a DNA molecule comprising an isolated or purified integrase-DNA binding moiety fusion protein gene refers to a DNA molecule including fusion protein coding sequences isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes genomic sequences, cDNA sequences or combinations thereof. "Isolated substantially away from other coding sequences" means that the gene of interest, in this case the fusion protein encoding gene, forms the significant part of the coding region of the DNA molecule, and that the DNA molecule does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

Another embodiment of the present invention is a purified nucleic acid molecule that encodes a protein in accordance with SEQ ID NOS:23, 25, 29, or 31, or a combination thereof, further defined as a recombinant vector. As used herein, the term "recombinant vector", refers to a vector that has been modified to contain a nucleic acid segment that encodes a fusion protein of the present invention, or fragment of interest thereof. The recombinant vector may be further defined as an expression vector comprising a promoter operatively linked to said fusion protein encoding nucleic acid molecule. In particular embodiments, the recombinant vector comprises a nucleic acid sequence in accordance with SEQ ID NOS:22, 24, 28, 30, a combination or a biologically functional fragment thereof. By way of example and not limitation, vectors may be further defined as a pT7-7, pET, pBluescript, pCMV, pUC and derivatives thereof, pBS24Ub, pYes2, pAC360 SV40, adenoviral, retroviral, yeast plasmids, *Baculovirus* or *Vaccinia* virus vector. Preferably, the expression vector is pT7-7, pET, pBS24Ub, pYes2, or pAC360.

A further embodiment of the present invention is a host cell, made recombinant with a recombinant vector comprising an integrase-DNA binding moiety encoding

gene. The recombinant host cell may be a prokaryotic or a eukaryotic cell, or a helper cell. In a more preferred embodiment, the recombinant host cell is a eukaryotic cell. As used herein, the term "engineered" or "recombinant" cell is intended to refer to a cell into which a recombinant gene, such as a gene encoding an integrase-DNA binding moiety, has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced gene. Thus, engineered cells are cells having a gene or genes introduced through the hand of man. Recombinantly introduced genes will either be in the form of a cDNA gene (i.e., they will not contain introns), a copy of a genomic gene, or will include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene, or combinations thereof. Preferred host cells may be further defined as any cell derived from a human, such as a stem cell, hepatocyte, fibroblast, or muscle cell; established cell lines such as CEM, MT-2, MT-4, T293, Jurkat, H9, HeLa, a COS cell, *Saccharomyces cerevisiae*, or *Escherichia coli* cell.

15

A further aspect of the present invention is a method of integrating a donor DNA molecule at or near a specific site or region thereof on a target DNA molecule. The method comprises the steps of i) selecting a DNA binding protein domain having binding affinity for the specific site or region thereof on the target DNA molecule, ii) constructing a fusion protein having an N-terminal retroviral integrase catalytic domain and the DNA binding protein domain at a C-terminus, and iii) contacting the donor DNA molecule, the target DNA molecule and the fusion protein, wherein the fusion protein facilitates integration of the donor DNA molecule at or near the specific site or region thereof of the target DNA molecule. In one embodiment of the invention, the donor DNA molecule comprises a gene encoding an integrase-DNA binding moiety fusion protein, in particular, the donor DNA molecule may comprise HIV-1 viral DNA having an integrase gene replaced with a gene encoding an integrase-DNA binding moiety fusion protein. The contacting step may further comprise the steps of i) incubating the fusion protein with the target DNA molecule to form an incubate, and ii) contacting the incubate with the donor DNA molecule.

30

In this method, the target DNA is DNA containing a defective gene, or DNA containing an oncogene or other disease causing gene, or DNA having no genes but is suitable as an acceptor site for exogenous DNA. A preferred DNA binding domain has binding affinity for nucleotide sequences found in regions of DNA as mentioned above
5 for preferred target DNA.

In this method, the retroviral integrase catalytic domain may be integrase from human immunodeficiency virus type 1 or type 2, or feline immunodeficiency virus. The DNA binding domain protein may be the LexA binding protein, and the specific
10 site on the target nucleic acid may be the LexA binding sequence. The LexA nucleotide sequence may be CTGTATGAGCATACAG (SEQ ID NO:21).

A further embodiment of the present invention is a method of inactivating an oncogene by integrating a donor DNA molecule at or near the oncogene, or regulatory
15 regions thereof. The method comprises i) selecting a DNA binding protein domain having binding affinity for the oncogene or regulatory regions thereof, ii) constructing a fusion protein having an N-terminal retroviral integrase catalytic domain and the DNA binding protein domain at a C-terminus, and iii) contacting a donor DNA molecule, the oncogene or regulatory regions thereof, and the fusion protein, wherein the fusion
20 protein facilitates integration of the donor DNA molecule at or near the oncogene or regulatory regions thereof, thereby inactivating the oncogene.

A further aspect of the present invention is a fusion protein comprising a catalytic domain of retroviral integrase and an N-terminal zinc finger domain having
25 binding specificity for a DNA molecule. In this case, the zinc finger domain is other than a zinc finger domain naturally occurring with the catalytic domain in a retroviral integrase molecule.

A fusion protein comprising an integrase catalytic domain fused to a protein
30 domain having affinity for a transcription factor is also an embodiment of the present invention. The transcription factor may be RNA polymerase III or TFIIC. The protein

domain having affinity for a transcription factor may be transcription factor IIIB-related factor (BRF).

A protein-oligonucleotide construct comprising an integrase catalytic domain covalently bonded to an oligonucleotide is also as aspect of the present invention.

Following long-standing patent law convention, the terms "a" and "an" mean "one or more" when used in this application, including the claims.

10

ABBREVIATIONS

IN	-	integrase
LA	-	LexA DNA binding protein
LABD	-	LexA DNA binding protein domain, from about amino acids 1-87 of LexA
WT	-	wild-type

15

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1. Formation of recombination intermediate. The initially blunt-ended linear viral DNA is cleaved by integrase, resulting in 3' ends recessed by 2 bases. The target DNA is cleaved with a 5-bp stagger, and the resulting 5'-P ends are joined to the 3'-OH ends of the viral DNA. The DNA joining reaction that gives rise to this recombination intermediate is referred to as integration (signified by a solid arrow) and to the reverse reaction that resolves its viral and target components as disintegration (signified by a broken arrow). Arrowheads indicate site of cleavage or strand exchange. The 3'-OH ends of DNA strands are denoted by half-arrows.

25
30

FIG. 2. DNA sequence and structure of Y-oligomer. The Y-oligomer substrate, which resembles the initial recombination intermediate shown in FIG. 1, was formed by annealing the following four oligonucleotides: T1, 16-mer; T3, 30-mer; V2, 21-mer; and the hybrid strand, V1.T2, 33-mer (SEQ ID NOS:12-15, respectively)

5

FIG. 3. Strand breakage and joining mediated by fusion proteins of the present invention. Schematic illustration of the expected products after disintegration of the Y-oligomer. Thick lines represent viral DNA sequences, and thin lines represent target DNA sequences. Closed circles denote the ³²P-labeled 5' ends. The length in nucleotides of each strand is indicated.

10

FIG. 4. Primary structures of HIV-1 integrase-*E. coli* LexA fusion proteins. Open and stippled boxes represent peptides derived from HIV-1 integrase and LexA proteins, respectively. Filled boxes represent the seven consecutive histidine residues (7xHis) used for protein purification. The left and right ends of the boxes denote the amino- and carboxy-terminus of the fusion proteins, respectively. The numbers in the boxes correspond to the amino acid residues from the native protein included in each fusion protein. Full-length HIV-1 integrase and LexA have 288 and 202 amino acids, respectively. LexA, full-length LexA protein; LexA BD, DNA-binding domain (amino acid residues 1-87) of LexA.

15

20

FIG. 5. DNA substrate for assaying distribution of integration sites. The LexA-binding sequence (underlined) was cloned into the *Kpn* I site of a plasmid derived from pBluescript KSII+. The resulting plasmid pBS-LA was digested with *Mbo* II to produce 6 fragments of different sizes (978, 639, 543, 409, 228, and 187 bp). The LexA-binding site is present in the 543-bp fragment. The arrows represent the primers used in PCR amplification of the integration products occurring in the plus or minus strand of the plasmid DNA. Primer BS+ is complementary to the plus strand of pBS-LA, whereas primer BS- is complementary to the minus strand. The numbers in parentheses denote the map positions of the sites for primer annealing and restriction enzyme cleavage. M, *Mbo* II.

25

30

FIG. 6. Nucleotide sequence (SEQ ID NO:22) and amino acid sequence (SEQ ID NO:23) of IN50-212/LABD, the HIV integrase catalytic domain (amino acids 50-212 of integrase) fused to the LexA DNA binding domain (amino acids 2-87 of LexA repressor). A peptide linker indicated by arrows (1) is the result of cloning techniques.

5

FIG. 7. Nucleotide sequence (SEQ ID NO:24) and amino acid sequence (SEQ ID NO:25) of IN1-288/LexA, the full-length HIV integrase (amino acids 1-288 of integrase) fused to the full-length LexA repressor (amino acids 2-202 of LexA repressor). A peptide linker indicated by arrows (1) is the result of cloning techniques.

10

FIG. 8. Full-length nucleotide sequence (SEQ ID NO:28), and full-length amino acid sequence (SEQ ID NO:29), of F-INI-281/LexA (full-length FIV integrase fused to full-length LexA repressor).

15

FIG. 9. Nucleotide sequence (SEQ ID NO:30) and amino acid sequence (SEQ ID NO:31) of F-INI-235/LexA (C-terminal truncated FIV integrase fused to full-length LexA repressor).

20

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention demonstrates that selection of sites in a target DNA molecule can be manipulated by fusing retroviral integrase with a sequence-specific DNA binding protein. A hybrid protein was constructed that has the *E. coli* LexA protein fused to the C-terminus of the HIV-1 integrase. The fusion protein, IN1-288/LA, retained the catalytic activities *in vitro* of the wild-type HIV-1 integrase (WT IN). Using an *in vitro* integration assay that included multiple DNA fragments as target DNA, IN1-288/LA preferentially integrated viral DNA into the fragment containing a DNA sequence specifically bound by LexA protein. No bias was observed when the LexA-binding sequence was absent, when the fusion protein was replaced by WT IN, or when LexA protein was added in the reaction containing IN1-288/LA. A

30

majority of the integration events mediated by IN1-288/LA occurred within 30 base pairs of DNA flanking the LexA-binding sequence.

5 The specificity toward LexA-binding sequence and the distribution and frequency of target site usage were unchanged when the integrase component of the fusion protein was replaced with a variant containing a truncation at the N- or C-terminus or both, suggesting that the domain involved in target site selection resides in the central core region of integrase. The integration bias observed with the integrase-LexA hybrid shows that one effective means of altering the selection of DNA
10 sites for integration is by fusing integrase to a sequence-specific DNA binding protein.

Two major improvements are a result of the targeted integration; i) safety, due to specific insertion that is targeted away from potentially harmful proto-oncogenes, and ii) improved expression, due to insertion that is targeted to cellular DNA regions that
15 are known for efficient and stable expression of genes.

Analysis of the distribution and frequency of integration sites indicates that the fusion proteins first bind specifically to the LexA-binding sequence and then mediate integration in the nearby regions flanking the binding site. The following observations
20 support this mechanism of action: (i) The preferred integration of the fusion proteins depended on the presence of LexA protein component, and was proportional to the binding affinities of the fusion proteins to the LexA-binding sequence. No preferred integration was observed with the wild-type or truncated HIV-1 integrases. (ii) The preferred integration depended on the presence of the LexA-binding sequence. In the
25 absence of the LexA-binding sequence in target DNA, the usage of target sites of fusion proteins was random and was identical to that of the wild-type integrase. In addition, preincubation of the target DNA with the fusion protein increased the integration specificity. (iii) The preferred integration was unique to the fusion proteins and no preferred integration was observed when the reaction was performed with a mixture of
30 wild-type integrase and LexA protein.

In certain embodiments, the invention concerns isolated DNA molecules and recombinant vectors which encode a fusion protein or peptide that includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NO:23, 25, 29, 31, a combination thereof or a biologically functional fragment thereof.

5 Naturally, where the DNA segment or vector encodes a full length integrase-LexA binding protein, or is intended for use in expressing the integrase-LexA binding protein, the most preferred sequences are those which are essentially as set forth in SEQ ID NO:25.

10 In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:22, 24, 28, 30, a combination thereof, or a biologically functional fragment thereof. The term "essentially as set forth in SEQ ID NO:22, 24, 28 or 30", is used in the same sense as described above and means that the

15 nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:22, 24, 28 or 30, and has relatively few codons which are not identical, or functionally equivalent, to the codons of SEQ ID NO:22, 24, 28 or 30. The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, as set forth in Table 1, and also refers to codons that

20 encode biologically equivalent amino acids.

Table 1

CODON DEGENERACY

<u>Amino Acids</u>		<u>Codons</u>					
Alanine	Ala	A	GCA	GCC	GCG	GCU	
Cysteine	Cys	C	UGC	UGU			
Aspartic acid	Asp	D	GAC	GAU			
Glutamic acid	Glu	E	GAA	GAG			
Phenylalanine	Phe	F	UUC	UUU			
Glycine	Gly	G	GGA	GGC	GGG	GGU	
Histidine	His	H	CAC	CAU			
Isoleucine	Ile	I	AUA	AUC	AUU		
Lysine	Lys	K	AAA	AAG			
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG
Methionine	Met	M	AUG				CUU
Asparagine	Asn	N	AAC	AAU			
Proline	Pro	P	CCA	CCC	CCG	CCU	
Glutamine	Gln	Q	CAA	CAG			
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG
							UCU

5

10

15

21

Threonine	Thr	T	ACA	ACC	ACG	ACU
Valine	Val	V	GUA	GUC	GUG	GUU
Tryptophan	Trp	W	UGG			
Tyrosine	Tyr	Y	UAC	UAU		

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences which may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, i.e., amino acids that form the junction between the integrase catalytic domain and the DNA binding protein domain of the fusion protein. The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences which have between about 70% and about 80%; or more preferably, between about 80% and about 90%; or even more preferably, between about 90% and about 99%; of nucleotides which are identical to the nucleotides of SEQ ID NO:22, 24, 28 or 30, will be sequences which are "essentially as set forth in SEQ ID NO:22, 24, 28 or 30". Sequences which are essentially the same as those set forth in SEQ ID NO:22, 24, 28 or 30 may also be functionally defined as sequences which are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:22, 24, 28 or 30 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art and are clearly set forth herein, for example conditions for use with PCR, and as described in the examples.

The present invention includes a purified nucleic acid molecule complementary, or essentially complementary, to the nucleic acid molecule having the sequence set

forth in SEQ ID NO:22, 24, 28 or 30. Nucleic acid sequences which are "complementary" are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences which are substantially complementary, as
5 may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:22, 24, 28 or 30 under relatively stringent conditions such as those described herein in the detailed description of the preferred embodiments. Complementary nucleotide sequences are useful for detection and purification of hybridizing nucleic acid molecules.

10

The present fusion proteins have an N-terminal histidine tag for purposes of facilitating purification of the fusion proteins. However, other molecular tags known to those of skill in the art may also be used in conjunction with the practise of the present invention. The present inventors also envision the preparation of further fusion
15 proteins and peptides, e.g., where the DNA binding moiety is from different DNA binding proteins as cited above, also where the fusion protein coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for further purification or immunodetection purposes (e.g., proteins which may be purified by affinity chromatography and enzyme label coding regions,
20 respectively).

The fusion proteins of the present invention have been successfully expressed in a prokaryotic expression system by the present inventors, especially using the pT7-7(His) vector in *E. coli* cells. Other expression systems contemplated by the present
25 inventors include, e.g., baculovirus-based, yeast-based, mammalian cell-based, or the like. For expression in this manner, one would position the coding sequences adjacent to and under the control of the promoter. It is understood in the art that to bring a coding sequence under the control of such a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame of the protein between
30 about 1 and about 50 nucleotides "downstream" of (i.e., 3' of) the chosen promoter.

Where eukaryotic expression is contemplated, one will also typically desire to incorporate into the transcriptional unit which includes the fusion protein gene, an appropriate polyadenylation site if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides
5 "downstream" of the termination site of the protein at a position prior to transcription termination.

It is contemplated that virtually any of the commonly employed host cells can be used in connection with the expression of the fusion proteins of the present invention
10 in accordance herewith. Examples include cell lines typically employed for eukaryotic expression such as COS, CV-1, CHO, murine fibroblasts C127 and 3T3, HeLa, HeLa S3, BS-C-1, HuTK-143B, or *Saccharomyces cerevisiae*.

Replication-defective, pseudotype viruses (a virus that cannot replicate on its
15 own, but needs complementary functions from a helper cell) and helper cells containing nucleic acids that encode a fusion protein of the present invention are an aspect of the invention. A pseudotype virus is made using two components, i) donor DNA having viral LTR-like ends, and ii) a helper cell encoding a fusion protein of the present invention and other essential viral proteins, and having necessary cellular machinery for
20 making virus. Donor DNA includes a packaging signal that allows the packaging of RNA made from donor DNA. This RNA together with viral proteins synthesized by the helper cell produce infectious virus. The virus is harvested and used to infect cells that are needing treatment. Alternatively, one could infect cells needing treatment with two vector constructs, one with donor DNA, and one with the retrovirus genome
25 carrying a fusion protein gene (but without the packaging signal).

Oligonucleotide sequences based on the fusion proteins of the present invention may be used as primers in a polymerase chain reaction or as hybridization probes to screen for the incorporation of fusion protein encoding sequences into a subject of
30 interest, a helper cell, for example.

DNA probes and primers useful in hybridization studies and PCR reactions may be derived from any portion of SEQ ID NO:22, 24, 28 or 30, and are generally at least about seventeen nucleotides in length. Therefore, probes and primers are specifically contemplated that comprise nucleotides 1 to 17, or 2 to 18, or 3 to 19 and so forth up to a probe comprising the last 17 nucleotides of the nucleotide sequence of SEQ ID NO:22, 24, 28 or 30. Thus, each probe would comprise at least about 17 linear nucleotides of the nucleotide sequence of SEQ ID NO:22, 24, 28 or 30, designated by the formula "n to n + 16," where n is an integer from 1 to about 753 or 1473, respectively. Longer probes that hybridize to the fusion protein gene under low, medium, medium-high and high stringency conditions are also contemplated, including those that comprise the entire nucleotide sequence of SEQ ID NO:22, 24, 28 or 30. Selected oligonucleotide subportions of the gene encoding a fusion protein of the present invention have significant utility as hybridization probes. Such probes may be used in the identification of genes encoding a fusion protein of the present invention that have been incorporated into helper cells or into a virus, for example. A general method for preparing oligonucleotides of various lengths and sequences is described by Caracciolo *et al.* (1989).

Preferred oligonucleotides resistant to *in vivo* hydrolysis may contain a phosphorothioate substitution at each base. Oligodeoxynucleotides or their phosphorothioate analogues may be synthesized using an Applied Biosystem 380B DNA synthesizer (Applied Biosystems, Inc., Foster City, CA).

A further embodiment of the invention is a purified nucleic acid molecule having at least a 17, 20, 25, 30, 50, 100, 200, 500, or 1000 nucleotide sequence that corresponds to, or is capable of hybridizing to the nucleic acid sequence of SEQ ID NO:22, 24, 28 or 30 under conditions standard for hybridization fidelity and stability. Furthermore, it is contemplated that nucleic acid molecules having a nucleotide sequence of SEQ ID NO:22, 24, 28 or 30 for stretches of between about 10 nucleotides to about 20 or to about 30 nucleotides will find particular utility, with even longer sequences, e.g., 40, 50, 150, 250, 450, even up to full length, being more preferred for

certain embodiments. These probes will be useful in hybridization embodiments, such as Southern and Northern blotting. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the complementary region may be varied, such as between about 20 and about 40 nucleotides, or even up to the full length of the nucleic acid as shown in SEQ ID NOS: 1, 9-13, 26 and 27 according to the complementary sequences one wishes to detect.

The use of a hybridization probe of about 10 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 10 bases in length are preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 20 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR technology of U.S. Patent 4,683,202 (herein incorporated by reference) or by introducing selected sequences into recombinant vectors for recombinant production.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In some embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or

spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

5 In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular
10 circumstances based on the particular criteria required (depending, for example, on the G+C contents, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label.

15 It will be understood that this invention is not limited to the particular nucleic acid and amino acid sequences having sequence identifiers as listed in Table 2. Therefore, DNA segments prepared in accordance with the present invention may also encode biologically functional equivalent proteins or peptides which have variant amino
20 acid sequences. Such sequences may arise as a consequence of codon redundancy and functional equivalency which are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be constructed via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on
25 considerations of the properties of the amino acids being exchanged.

Table 2 lists the identity of sequences of the present disclosure having sequence identifiers.

Table 2
Identification of Sequences Having Sequence Identifiers

	SEQ ID NO:	IDENTITY
5	1	5'-GAAGGAGATATACATATGTTTTAGATGGA-3', primer for the N-terminus of the full-length integrase
	2	5'-TAGACTCATATGCATGGACAAGTA-3', primer for the N-terminus of the N-terminally truncated (amino acid residues 1-50) integrase
	3	5'-GCTAGAGGTACCATCCTCATCCTGTCTACT-3', primer for the C terminus of the full-length integrase
	4	5'-GCTAGAGGTACCAACTGGATCTCTGCTGTC-3', primer for the C terminus of the C-terminally truncated (amino acid residues 235-288) integrase
	5	5'-CAGTCAGGTACCAAAGCGTTAACGGCCAGG-3', primer for the N terminus of the <i>lexA</i> gene
10	6	5'-ATAGGATCC TT ACAGCCAGTCGCCGTTGCG-3', primer for the C terminus of the full-length LexA protein
	7	5'-ATTGGATCC TT ATGGTTCACCGGCAGC-3', primer for the C terminus of the DNA-binding domain (amino acids 1 to 87) of LexA protein
	8	5'-TAATG <u>CATCACCATCACCATC</u> ACCA-3', double stranded oligonucleotide allowed insertion of ATG initiation codon (italicized) and seven histidine codons (underlined) into the unique <i>Nde</i> I site of pT7-7
	9	5'-TATGGTGATGGTGATGGTGATGCAT-3', complement of SEQ ID NO:8 with added nucleotides
	10	5'-CAGGC <u>CTGTATGAGCATA</u> CAGGTAC-3', double stranded oligonucleotide allowed preparation of a plasmid that contains a single specific binding site for LexA protein
15	11	5'- <u>CTGTATGCTCATA</u> CAGGCCTGGTAC-3', complement to SEQ ID NO:10 with nucleotides added
	12	T1 substrate for integration assay, 5'-CAGCAACGCAAGCTTG-3'
	13	T3 substrate for integration assay, 5'-GTCGACCTGCAGCCCAAGCTTGCGTTGCTG-3'
	14	V2 substrate for integration assay, 5'-ACTGCTAGAGATTTTCCACAT-3'

- 15 V1/T2 substrate for integration assay,
5'-ATGTGGAAAATCTCTAGCAGGCTGCAGGTCGAC-3'
- 16 C220 substrate for integration assay,
5'-ATGTGGAAAATCTCTAGCAGT-3',
- 17 B2-1 substrate for integration assay,
5'-ATGTGGAAAATCTCTAGCA-3'
- 18 5'-CATTAATGCAGCTGGCACGA-3', BS+ PCR primer for analysis
of the integration events occurring in the plus strand of plasmid DNA
- 5 19 5'-TAATACGACTCACTATAGGG-3', BS- PCR primer for analysis of
the integration events occurring in the minus strand
- 20 CTGTNNNNNNNACAG, LexA consensus binding sequence
- 21 CTGTATGAGCATACAG, LexA binding sequence
- 22 Nucleotide sequence of IN50-212/LABD
- 23 Amino acid sequence of IN50-212/LABD
- 10 24 Nucleotide sequence of IN1-288/LexA
- 25 Amino acid sequence of IN1-288/LexA
- 26 A 5'-3' oligonucleotide primer for FIV integrase,
5'-CCAGTGCATATGTCCTCTTGGGTTGACAGA-3'
- 27 A 5'-3' oligonucleotide primer for FIV integrase,
5'-CAGTCAGGTACCCTCATCCCCTTCAGG-3'
- 28 Nucleotide sequence of F-INI-281/Lex A (full-length FIV integrase fused to full-
length LexA repressor) (Figure 8)
- 15 29 Amino acid sequence of F-INI-281/Lex A (full-length FIV integrase fused to full-
length LexA repressor) (Figure 8)
- 30 Nucleotide sequence of F-INI-235/LexA (C-terminal truncated FIV integrase fused
to full-length LexA repressor) (Figure 9)
- 31 Amino acid sequence of F-INI-235/Lex A (C-terminal truncated FIV integrase
fused to full-length LexA repressor) (Figure 9)
- 32 Nucleic acid sequence, a 3' primer for FIV INI-235,
5'-GCTAGAGGTACCTTTCTTATCTTTTGATC
- 33 A 5' primer for the *rlet* gene,
5'-CAGTCAGGTACCTCTAGATTAGATAAAAGT-3'
- 20 34 A 3' primer for the *rlet* gene,
5'-CAGTCAGGATCCGGACCCACTTTCACATTT-3'

In some aspects, the present invention provides a purified integrase-DNA binding moiety fusion protein having an amino acid sequence essentially as set forth in SEQ ID NO:23, 25, 29 or 31. Peptides of a fusion protein are useful for designing oligonucleotides for screening for the presence of the gene encoding said fusion protein. Peptides having less than about 45 amino acid residues may be chemically synthesized by the solid phase method of Merrifield (1963) in light of this disclosure. The Merrifield reference is specifically incorporated by reference herein, using an automatic peptide synthesizer with standard t-butoxycarbonyl (t-Boc) chemistry that is well known to one skilled in this art. The amino acid composition of the synthesized peptides may be determined by amino acid analysis with an automated amino acid analyzer to confirm that they correspond to the expected compositions. The purity of the peptides may be determined by sequence analysis or HPLC.

In still another embodiment of the present invention, methods of preparing an integrase-DNA binding moiety protein composition are provided. In one aspect, the method comprises growing recombinant host cells comprising a vector that encodes a protein which includes an amino acid sequence in accordance with SEQ ID NO:23, 25, 29 or 31, under conditions permitting nucleic acid expression and protein production followed by recovering the protein so produced. The host cell, conditions permitting nucleic acid expression, protein production and recovery, will be known to those of skill in the art, in light of the present disclosure of the fusion proteins of the invention. A preferred host cell is an *E. coli* cell.

Modifications and changes may be made in the sequence of the fusion proteins of the present invention and still obtain a peptide or protein having like or otherwise desirable characteristics. For example, certain amino acids may be substituted for other amino acids in a peptide without appreciable loss of function. Since it is the interactive capacity and nature of an amino acid sequence that defines the peptide's functional activity, certain amino acid sequences may be chosen (or, of course, its underlying DNA coding sequence) and nevertheless obtain a peptide with like properties. It is thus contemplated by the inventors that certain changes may be made in the sequence of an

integrase-DNA binding moiety fusion protein (or underlying DNA) without appreciable loss of its ability to function.

Substitution of like amino acids can be made on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 \pm 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent peptide. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are more preferred, and those within ± 0.5 are most preferred.

15

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

20

Two designations for amino acids are used interchangeably throughout this application, as is common practice in the art. Alanine = Ala (A); Arginine = Arg (R); Aspartate = Asp (D); Asparagine = Asn (N); Cysteine = Cys (C); Glutamate = Glu (E); Glutamine = Gln (Q); Glycine = Gly (G); Histidine = His (H); Isoleucine = Ile (I); Leucine = Leu (L); Lysine = Lys (K); Methionine = Met (M); Phenylalanine = Phe (F); Proline = Pro (P); Serine = Ser (S); Threonine = Thr (T); Tryptophan = Trp (W); Tyrosine = Tyr (Y); Valine = Val (V).

30

While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes may be effected by alteration of the encoding DNA, taking into consideration also that the genetic code is degenerate and that two or more codons may code for the same amino acid.

5

Another aspect of the present invention provides therapeutic agents for the incorporation of a therapeutic gene or for the inactivation of an oncogene, for example, in an animal. The therapeutic agent comprises an admixture of integrase-DNA binding moiety fusion protein in a pharmaceutically acceptable excipient. Most preferably, the therapeutic agent will be formulated so as to be suitable for injection. Pharmacologically active fusion proteins may also be provided to a subject via gene therapy. Many different vehicles exist for accomplishing this end, such as incorporation of the fusion protein gene, or fragment thereof, into an adenovirus, retrovirus, or other techniques known to those of skill in the art in light of the present disclosure. *Ex vivo* gene therapy is also contemplated as another mode of administration.

10
15

Such preparations should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 60% of the weight of the unit. The amount of active compounds in such therapeutically useful compositions is such that a suitable dosage will be obtained.

20

The active compounds may be administered parenterally or intraperitoneally. Solutions of the active compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

25

30

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for

pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. See, for example, Remington (1995), which
5 reference is incorporated by reference herein.

In another aspect, the present invention includes an antibody that is immunoreactive with an integrase-DNA binding moiety fusion polypeptide as described for the invention. An antibody can be a polyclonal or a monoclonal antibody. In some
10 embodiments, the antibody is a monoclonal antibody. Means for preparing and characterizing antibodies are well known in the art (See, e.g., *Antibodies "A Laboratory Manual*, E. Howell and D. Lane, Cold Spring Harbor Laboratory, 1988).

The present invention in still another aspect defines an immunoassay for the
15 detection of an integrase-DNA binding moiety fusion protein in a biological sample. In one particular embodiment of the immunoassay, the immunoassay comprises; preparing an antibody having binding specificity for the fusion protein to provide an anti-fusion protein antibody, incubating the anti-fusion protein antibody with the biological sample for a sufficient time to permit binding between antibody and fusion
20 protein present in said biological sample, and determining the presence of bound antibody by contacting the incubate of the sample and antibody with a detectably labeled antibody specific for the anti-fusion protein antibody, wherein the presence of anti-fusion protein antibody in the biological sample is detectable as the measure of the detectably labeled antibody from the biological sample.

25

By way of example, the antibody may be labeled with any of a variety of detectable molecular labeling tags. Such include, an enzyme-linked antibody, a fluorescent-tagged antibody, or a radio-labelled antibody.

30

Even though the invention has been described with a certain degree of particularity, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art in light of the foregoing disclosure. Accordingly, it is intended that all such alternatives, modifications, and variations which fall within the spirit and the scope of the invention be embraced by the defined claims.

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

Primary Structures of Integrase-LexA Fusion Proteins

The present example provides constructs of fusion proteins studied as part of the present invention.

The selection of integration sites was studied by fusing integrase to the *E.coli* LexA repressor, a sequence-specific DNA binding protein. The LexA repressor of *E.coli* negatively regulates the transcription of about 20 SOS genes that are mostly involved in DNA repair, mutagenesis, DNA replication, and cell division (for reviews, see Little and Mount, 1982; and Schnarr *et al.*, 1991). LexA protein contains two domains: the first 87 amino acids at the N-terminus constitute the DNA binding domain, and amino acid residues 88 to 202 constitute the dimerization domain (Fogh *et al.*, 1994; Schnarr *et al.*, 1988; Thliveris and Mount, 1992). LexA protein binds specifically to a 16-bp DNA sequence that consists of two dyad symmetric half-sites of 8 bp each, starting with a highly conserved CTG trinucleotide and followed by a less

conserved but AT-rich 5-bp sequence (Wertman and Mount, 1985). The sequence used in this study corresponds to the *recA* operator, a site that LexA binds with high affinity (Lewis *et al.*, 1994). The ability of LexA to bind to specific DNA sequences is retained after LexA is fused to various other proteins (Brent and Ptashne, 1985; Golemis and Brent, 1992; Schmidt-Dorr *et al.*, 1991; Wang and Stillman, 1993).

HIV-1 integrase and the *lexA* genes were obtained from plasmids pT7-7-IN (Vincent *et al.*, 1993) and pBTM117, respectively. A parent plasmid to pBTM117, pBTM116, is described in Vojtek (1993). For purposes of the present invention, these plasmids are essentially the same. The genes were amplified by polymerase chain reaction (PCR). Oligonucleotide primers used in PCR were from Operon Technologies, Inc. (Alameda, CA). The primers for the N-terminus of the full-length and the N-terminus truncated (amino acid residues 1-50) integrases were 5'-GAAGGAGATATACATATGTTTTAGATGGA-3' (SEQ ID NO:1) and 5'-TAGACTCATATGCATGGACAAGTA-3' (SEQ ID NO:2), respectively. The N-terminus primers contain an *Nde* I site. The primers for the C terminus of the full-length and the C-terminus truncated (amino acid residues 235-288) integrases were 5'-GCTAGAGGTACCATCCTCATCCTGTCTACT-3' (SEQ ID NO:3) and 5'-GCTAGAGGTACCAACTGGATCTCTGCTGTC-3', (SEQ ID NO:4) respectively. The C-terminus primers contain a *Kpn* I site.

The primer for the N terminus of the *lexA* gene was 5'-CAGTCAGGTACCAAAGCGTTAACGGCCAGG-3' (SEQ ID NO:5) and contains a *Kpn* I site. The primers for the C terminus of the full-length and the DNA-binding domain (amino acids 1 to 87) of LexA protein were 5'-ATAGGATCCTTACAGCCAGTCGCCGTTGCG-3' (SEQ ID NO:6) and 5'-ATTGGATCCTTATGGTTCACCGGCAGC-3' (SEQ ID NO:7), respectively. The C-terminus primers for the *lexA* gene contain a *Bam*H I site and a stop codon (italicized). After PCR, the DNA fragments containing the integrase gene were cut with *Nde* I and *Kpn* I, and the DNA fragments containing the *lexA* gene were cut with *Kpn* I and *Bam*H I. The cleaved DNA fragments were purified with the Qiaex gel extraction

kit (Qiagen) and ligated to pT7-7(His) plasmid DNA, previously cut with *Nde* I and *Bam*H I. The plasmid pT7-7(His) is derived from pT7-7, a T7 RNA polymerase-promoter system (Tabor and Richardson, 1985), and was prepared by inserting a double-stranded oligonucleotide
5 (5'-TAATGCATCACCATCACCATCACCA-3' (SEQ ID NO:8) and 5'-TATGGTGATGGTGATGGTGATGCAT-3', (SEQ ID NO:9)) that contains an ATG initiation codon (italicized) and seven histidine codons (underlined) into the unique *Nde* I site of pT7-7.

10 To prepare a plasmid that contains a single specific binding site for LexA protein, a double-stranded oligonucleotide (5'-CAGGCCTGTATGAGCATAAGGTAC-3', (SEQ ID NO:10) and 5'-CTGTATGCTCATAAGGCCTGGTAC-3', (SEQ ID NO:11)) containing the *recA* operator sequence (underlined) was inserted into the *Kpn* I site of a plasmid derived
15 from pBluescript KSII+ (Stratagene), resulting in pBS-LA (FIG. 5).

Standard cloning procedures were followed (Sambrook, *et al.*, 1989). The sequences of all the PCR-amplified DNA fragments were verified by restriction analysis and the dideoxynucleotide chain termination method. Sequencing reactions
20 were carried out with a modified T7 polymerase (Sequenase version 2.0, U.S.Biochemicals, Cleveland, OH) according to manufacturer's specification.

The various fusion proteins constructed and studied in this report are shown in FIG. 4. The fusion protein consisting of full-length HIV-1 integrase fused to LexA
25 (IN1-288/LA) serves as the prototype. Two fusion constructs, IN1-288/LABD and IN1-234/LABD, were prepared for determining whether fusion proteins containing only the DNA binding domain of LexA was sufficient for altering target site selection. Since the central core of integrase contains the catalytic site and the C-terminus of integrase shows non-specific DNA binding (Engelman *et al.*, 1994; Schauer and Billich, 1992; Vink *et al.*, 1993; Woerner *et al.*, 1992), several fusion constructs were prepared that
30 include various truncated forms of integrase, such as IN1-234/LA, IN50-288/LA, and

IN50-234/LA. These constructs would indicate whether the fusion proteins containing truncated integrase, when compared with those containing full-length integrase, have an increased specificity toward LexA-binding sequence in target site usage.

5

EXAMPLE 2

In vitro Activities of the Purified Fusion Proteins

The present example provides studies carried out to demonstrate 3'-end processing and 3'-end joining activities, and footprinting analyses of protein binding to a LexA-recognition sequence.

Expression and purification of the fusion proteins. The DNA constructs were transformed into *E. coli* BL21 (DE3). The cells were grown at 30°C. When the OD₆₀₀ was 0.8-1, 0.4 mM isopropyl-1-thio-β-D-galactopyranoside was added for expression induction, and the culture was grown for an additional 3 hours.

Purification in denaturing conditions. The cell pellet was resuspended in a buffer (5 ml buffer per gram of cells) containing 20 mM Tris-HCl, pH 8, 0.5 M NaCl and 6 M guanidine-HCl (Buffer A). The suspension was frozen and thawed, homogenized by stirring for one hour at room temperature, and spun at 27,000 x g for 30 min at 4°C. The supernatant was passed twice over a Ni²⁺-charged metal-chelating column (Qiagen) in the presence of 6M guanidine-HCl at room temperature. Each column passage included a wash with Buffer A, a second wash with Buffer A plus 20 mM imidazole, and elution with a linear gradient from Buffer A plus 20 mM imidazole to Buffer A plus 500 mM imidazole. The fractions containing the protein were pooled and dialyzed in a stepwise manner against buffer B (25mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid [HEPES, pH 7.5], 1 mM EDTA, 10 mM dithiothreitol [DTT], 300 mM NaCl, 10% glycerol, 10 mM 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate [CHAPS]) plus 1M guanidine-HCl at 4°C. A 1.5-ml protein sample was then applied at 0.5 ml/min to a Superdex 75 (Pharmacia Biotech)

column (about 100-ml resin bed volume) at 4°C. The fractions containing the protein were pooled and dialyzed against Buffer B.

Purification in native conditions. The cell pellet was resuspended in a buffer containing a final concentration of 20 mM HEPES, pH 7.5, 1 M NaCl, 10% glycerol, 5 mM 2-mercaptoethanol, 0.2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.2 mg/ml lysozyme, and 0.1% Nonidet P-40. The cell suspension was sonicated and centrifuged at 100,000 x g for 1 h at 4°C. The supernatant, after dialysis against buffer C (20 mM HEPES, pH 7.5, 1 M NaCl, 10% glycerol, 5 mM 2-mercaptoethanol, 0.1% Nonidet P-40), was incubated on ice for 2 hours with the Ni-NTA resin. The resin was sequentially washed with buffer C, buffer C plus 10 mM imidazole, buffer C plus 50 mM imidazole, and buffer C plus 70 mM imidazole. The resin was then packed in a column and the protein was eluted with a linear gradient from buffer C plus 70 mM imidazole to buffer C plus 500 mM imidazole. The fractions containing the protein were pooled, concentrated on a Centricon-10 column (Amicon), and dialyzed against the final buffer (20 mM HEPES, pH 7.5, 0.5 M NaCl, 20% glycerol, 0.1 mM EDTA, 1 mM DTT and 10 mM CHAPS). Protein concentrations were determined by the Bradford method (Bio-Rad) using bovine serum albumin (BSA) as a standard.

20

The wild-type integrase and the fusion proteins IN1-234/LA and IN50-234/LA were purified in both native and denaturing conditions. For each protein, no difference in activity was observed when the protein was purified in either condition. The proteins IN50-234 and IN50-288/LA were purified under the native condition only, whereas the proteins IN1-234, IN1-288/LABD, and IN1-234/LABD were purified under the denaturing condition only. A Coomassie Blue-stained SDS-PAGE of various purified proteins indicated bands of the expected molecular weight for wild-type integrase, IN1-288/LABD, IN1-288/LA, wild type LexA, IN1-234, IN1-234/LA, and IN1-234/LABD. One microgram of each purified protein was run on a 12% SDS-PAGE. Molecular weight standards were from Gibco BRL (Grand Island, NY).

30

Footprinting analysis of DNA binding. The pBS-LA plasmid DNA, which contains the LexA-binding sequence, was digested with *Bam*HI. The linearized DNA was labeled at the 5' end with [γ -³²P] ATP using T4 polynucleotide kinase and digested with *Pvu* II. The 311-base pair (bp) singly end-labeled fragment containing the LexA-binding sequence was isolated from a 1.2% agarose gel with the Qiaex gel extraction kit (Qiagen, Chatsworth, CA). About 6 fmol (30,000 cpm) of the fragment was incubated with the protein at room temperature for 30 min, in a buffer containing a final concentration of 20 mM HEPES, pH 7.5, 10 mM DTT, 0.05% Nonidet P-40, 1.5 mM CaCl₂, 2.5 mM MgCl₂, 100 μ g/ml BSA, 2 μ g/ml poly dI-dC, and 50 mM NaCl. The samples were digested with 2 ng/ml DNase I for 3 min at room temperature. The digestion was stopped by the addition of 18 mM EDTA, and the samples were deproteinized by phenol-chloroform extraction, ethanol precipitated in the presence of 10 μ g of tRNA as a carrier, and resuspended in 5 μ l of formamide, 10 mM EDTA. After denaturation at 90°C for 3 min, the samples were analyzed by electrophoresis through a 5% denaturing polyacrylamide gel.

Integration assays. The 3'-end processing, 3'-end joining, and disintegration activities of the fusion proteins were assayed as previously described (Chow *et al.*, 1992; Vincent *et al.*, 1993).

The following oligonucleotides (Operon Technologies, Inc., Alameda, CA) were used as DNA substrates: T1 (16 mer), 5'-CAGCAACGCAAGCTTG-3', (SEQ ID NO:12); T3 (30 mer), 5'-GTCGACCTGCAGCCCAAGCTTGCGTTGCTG-3', (SEQ ID NO:13); V2 (21 mer), 5'-ACTGCTAGAGATTTTCCACAT-3', (SEQ ID NO:14); V1/T2 (33 mer), 5'-ATGTGGAAAATCTCTAGCAGGCTGCAGGTCGAC-3', (SEQ ID NO:15); C220 (21 mer), 5'-ATGTGGAAAATCTCTAGCAGT-3', (SEQ ID NO:16); B2-1 (19 mer), 5'-ATGTGGAAAATCTCTAGCA-3', (SEQ ID NO:17). The oligonucleotides were purified by electrophoresis through a 15% denaturing polyacrylamide gel. Oligonucleotides T1, C220 and B2-1 were labeled at the 5' end with [γ -³²P] ATP (6000 Ci/mmol, Amersham, Arlington Heights, IL) using T4 polynucleotide kinase.

The 3'-end processing and 3'-end joining substrate, which corresponds to the terminal 21 nucleotides of the U5 end of viral DNA, was prepared by annealing the labeled C220 strand with its complementary oligonucleotide V2. The preprocessed substrate, which resembles the viral U5 end after 3'-end processing, was prepared by
5 annealing the labeled B2-1 strand with the V2 strand and was used to assay only the 3'-end joining activity. A reaction was carried out with 5 nM of the U5 end oligonucleotide (C220/V2) and 100 nM of protein. The substrate was the 21-mer, and the 3'-end processing product was a 19-mer. Strand transfer products were visible on the gel also.

10

The substrate for assaying disintegration activity, the Y-oligomer, was prepared by annealing the labeled T1 strand with oligonucleotides T3, V2 and V1/T2 (Chow *et al.*, 1992). In a 20 μ l volume, the DNA substrate (0.1 pmol) was incubated with the protein for one hour at 37°C in the standard reaction buffer containing a final
15 concentration of 20 mM HEPES, pH 7.5, 10 mM DTT, 0.05% Nonidet P-40 and 10 mM MnCl₂. The reaction was stopped by the addition of 18 mM EDTA. The reaction products were heated at 90°C for 3 min before analysis by electrophoresis on 15% polyacrylamide gels with 7M urea in Tris-borate-EDTA buffer. A reaction was carried out with 5 nM of the Y-oligomer substrate and 250 nM of protein. The 5'-end-labeled
20 T1 strand of the Y-substrate migrated as a 16-nucleotide on the denaturing gel. The disintegration product was a 30-mer. Controls were done in the absence of protein.

In vitro activities of the purified fusion proteins. All fusion proteins were first tested using the oligonucleotide-based assays for their abilities to mediate 3'-end
25 processing, 3'-end joining, and disintegration. Results of autoradiographs are summarized in Table 3.

Table 3. Summary of *in vitro* activities^a of HIV-1 integrase mutants and fusion proteins

5	Integrase derivative	3'-End processing	3'-End joining	Disintegration
	INI-288/LA	++	++	+++
	INI-288/LABD	++	++	+++
	INI-234/LA	-	- ^b	+++
10	INI-234/LABD	-	- ^b	+++
	INI-234	-	- ^b	++
	IN50-288/LA	-	- ^b	++
	IN50-234/LA	-	- ^b	++
	IN50-234	-	- ^b	+

15

^a Relative activities are expressed as the percentage of the activity of wild-type HIV-1 integrase. +, 50% or less; ++, wild-type level of activity; +++, 150% or more; -, no activity.

20

^b Although little or no 3'-end joining activity was observed using the oligonucleotide-based assay, strand transfer products were detected using the PCR-based assay.

25

30

Fusing integrase with either full-length or only the DNA-binding domain of LexA did not change appreciably the catalytic activities of integrase, and the two fusion proteins, IN1-288/LA and IN1-288/LABD, showed similar 3'-end processing and 3'-end joining activities as did WT IN. For the 3'-end joining reaction, the patterns and the intensities of the recombinant products were similar among WT IN, IN1-288/LA, and IN1-288/LABD, indicating that fusion with LexA also did not alter the recognition by integrase of target DNA containing non-specific sequences.

Integrases containing various truncations, and fusion proteins containing truncated integrase were inactive in 3'-end joining and 3'-end processing but retained disintegration activity (Table 1). Although the truncated variants of integrase, either by themselves or fused with LexA, did not exhibit 3'-end joining activity using the oligonucleotide-based assays, the ability of these proteins to mediate 3'-end joining was demonstrated by a more sensitive PCR-based assay. IN1-186/LA did not display any catalytic activities. Fusing WT IN or truncated integrase to full length LexA or only the DNA-binding domain of LexA increased the disintegration activity of the cognate protein.

The abilities of the constructed fusion proteins to recognize and bind specifically to a LexA-binding sequence were examined by DNase I footprinting analysis. The control proteins, WT IN and IN1-234 did not display any specific DNA binding on this DNA fragment, and the gel banding patterns were identical to that obtained in the absence of any protein. With the wild-type LexA protein, a protected region of about 25 bp in size was observed. Protection of the LexA-binding sequence was also observed with the various fusion proteins IN1-288/LA, IN1-288/LABD, IN1-234/LA, and IN1-234/LABD; providing direct evidence for sequence-specific DNA binding of these proteins. By calculating the amount of protein necessary to protect 50% of the sequence (Brenowitz *et al.*, 1993; 1986), the dissociation constant (K_d) of the following proteins was estimated: LexA, 2nM; IN1-288/LA, 10 nM; IN1-288/LABD, 250 nM; IN1-234/LA, 5 nM and IN1-234/LABD, 150 nM. The stronger protection displayed by fusion proteins containing full-length LexA, when compared to those displayed by fusion proteins containing only the DNA binding domain of LexA, suggests that the full-length LexA protein fused to the HIV-1 integrase is still able to dimerize, which provides a cooperative mode of binding to the operator. For IN1-288/LA and IN1-234/LA, the size of protection was identical to that of wild-type LexA protein, suggesting that a LexA dimer component of the fusion protein is primarily responsible for DNA binding.

EXAMPLE 3
Integrase-LexA Fusion Proteins Direct
Selective Integration into DNA

5 The present example demonstrates selective integration into DNA mediated by integrase-LexA fusion proteins and the effect of preincubation of IN1-288/LA with target DNA.

10 *Assays for distribution of integration sites.* The donor DNA substrate used to assay the distribution of integration sites of the HIV integrase-LexA fusion proteins was the preprocessed U5 DNA substrate (B2-1/V2). The target DNA was the plasmid pBS-LA, as described in Example 1. The distribution of the integration sites was analyzed by the following assay and the PCR assay of Example 5.

15 *Agarose gel assay.* pBS-LA was cleaved with *Mbo* II to generate multiple fragments ranging in size from 0.1 to 1 kbp (see FIG. 5). The fragment that contains the LexA-binding sequence is 543 bp in length (FIG. 5). The DNA fragments (1 μ g) were incubated with WT IN or with the fusion protein for 5 min on ice in the standard reaction buffer. The integration reaction was started by adding 15 nM of the
20 preprocessed U5 substrate (B2-1/V2), labeled at the 5' end of B2-1, and transferring the reaction to 37°C. After a 30-min incubation, the reaction was stopped by adding 2 μ l of 0.2 M EDTA, pH 8.0. The total reaction volume was 20 μ l. The reaction product was mixed with a 1/6 volume of loading buffer (30% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol) and separated by electrophoresis on a 1.5% agarose gel in
25 Tris-borate-EDTA buffer. After electrophoresis, the DNA fragments were visualized by ethidium bromide staining (0.5 μ g/ml) and autoradiography.

30 *Directed integration mediated by integrase-LexA fusion protein.* Formation of recombinant products by integration of the labeled U5 DNA into target DNA was assayed by the appearance of labeled, high molecular weight DNA fragments. In the presence of WT IN (no fusion), integration appeared to be random and occurred in each

of the DNA fragments with similar frequency. The integration frequency using WT IN increased at higher protein concentrations but the relative intensity among the various DNA fragments remained the same. In contrast, integration of the U5 DNA by the fusion protein IN1-288/LA was unevenly distributed and showed a bias towards the DNA fragment containing the LexA-binding sequence. In the presence of 2 pmol fusion protein, the molar ratio between the DNA fragment containing the LexA-binding sequence and the IN1-288/LA dimer was about 1:1. The 543-bp lexA-containing DNA fragment was preferred approximately 14-50 fold over the other fragments. At higher concentrations of IN1-288/LA, the integration frequency increased but the bias became less apparent. In the reaction containing 10 pmol of IN1-288/LA, the preference for the 543-bp fragment was approximately 4-fold. The frequency of integration mediated by wild-type or IN1-288/LA into the two smallest *Mbo* II-cleaved products, 187 and 228, were approximately 3-fold less than that of the 409-bp fragment.

These results show that integration mediated by the integrase-LexA fusion protein was directed through specific DNA binding towards the fragment containing the LexA-binding sequence. The decrease in the selectivity at higher protein concentrations may be due to a saturation of binding of the LexA-binding site, which then caused the excess fusion protein to mediate integration randomly into other DNA fragments.

A similar study was carried out using IN1-288/LABD as the integration protein. The result obtained with IN1-288/LABD was similar to that obtained with IN1-288/LA. The distribution of integration sites of the fusion protein containing only the LexA-binding domain also exhibited a preference for the LexA-binding sequence but the bias was approximately two-fold less than that of IN1-288/LA. This result could be due to the lower binding affinity of IN1-288/LABD in comparison to IN1-288/LA, and is consistent with results showing that DNA binding by many LexA derivatives that contain the C-terminal dimerization domain is considerably higher than binding by fusions that lack it (Golemis and Brent, 1992).

Because of the poor 3'-end joining activity of the truncated integrase-LexA fusion proteins (Table 1), the distribution of their integration sites was not determined using the agarose gel assay. Instead, the target site usage of these fusion proteins was examined using a more sensitive PCR-based assay (Example 5).

5

Effect of preincubation of IN1-288/LA with target DNA. Two picomoles of WT IN or IN1-288/LA was preincubated with 1 μ g of *Mbo* II-cleaved pBS-LA at room temperature for 0, 1, 5, 10, 20, or 30 min before the addition of the preprocessed U5 DNA. In other tubes, the protein was preincubated at room temperature for 5 min with the preprocessed U5 DNA before the reaction was started by adding target DNA.

10

Results demonstrated that the target site selection was influenced by whether the fusion protein was preincubated with the target DNA or the donor DNA. The DNA fragment containing the LexA-binding sequence was preferred when the fusion protein was preincubated with the target DNA, although the time of preincubation was not critical. In contrast, when the fusion protein was preincubated with the donor DNA, the integration events became more evenly distributed. In the case of the wild-type protein, no difference was observed whether the protein was preincubated with the target or donor DNA. The result is consistent with the preferred integration being mediated by the specific interaction between the fusion protein and the LexA-binding sequence, and that such an interaction is promoted when the fusion protein is preincubated with the target DNA.

15

20

EXAMPLE 4

25

Directed Integration by the Fusion Protein Depends on LexA-Binding Site and can be Competed by LexA Protein

30

The present example confirms that integration by the fusion protein at a targeted site is directed by a DNA binding protein domain having binding specificity for a target nucleotide sequence, such as for example the presence of the LexA-binding sequence. The present inventor examined the distribution of integration sites into DNA fragments

generated from *Mbo* II cleavage of the parental plasmid pBS, which contains no LexA-binding sequence as a model.

Integration of preprocessed U5 DNA was carried out by WT IN or IN1-288/LA
5 using 1 μ g of *Mbo* II-cleaved pBS or *Mbo* II-cleaved pBS-LA as the target DNA. In
pBS, which has no LexA-binding sequence, the fragment corresponding to the 543-bp
fragment of pBS-LA is 521 bp in length. Under the identical reaction conditions and
in the absence of LexA-binding sequence in the target DNA, IN1-288/LA fusion protein
10 showed no bias in the frequency of integration. The result indicates that the 543-bp
fragment, except in the presence of the LexA-binding sequence, possessed no preferred
sequence or DNA features that could have caused the directed integration.

A competition experiment was carried out to test the hypothesis that the directed
integration observed with the fusion protein was mediated by its specific binding to the
15 LexA-binding sequence. Integration reactions were performed with 2 pmol WT IN or
IN1-288/LA in the presence of 0-20 pmol of LexA repressor. The LexA protein was
preincubated first with the target DNA (*Mbo* II-cleaved pBS-LA) for 5 min at room
temperature before the reaction was started by adding the WT IN or the IN1-288/LA
and 0.3 pmol of the 5'-end labeled U5 DNA. In the presence of an increasing amount
20 of LexA protein, the preferred integration mediated by IN1-288/LA into the DNA
fragment containing the LexA-binding sequence correspondingly diminished, and the
integration became more evenly distributed among all DNA fragments. The result is
consistent with the model that LexA protein competes with the fusion protein for the
LexA-binding site, resulting in 'free' fusion protein that mediates random integration.
25 Moreover, the LexA-bound DNA fragment, with the LexA-binding site being occupied,
can no longer be specifically targeted. As a negative control, addition of LexA protein
to the reaction containing WT IN had no effect on the distribution of integration sites.
The unaltered usage of integration sites by WT IN and LexA protein also ruled out the
possibility that the directed integration by the fusion protein could be an artifact
30 resulting from DNA distortion induced by LexA protein binding.

EXAMPLE 5
Detailed Analysis of Integration Sites
Using the PCR-Based Assay

5 The present example provides a detailed analysis of the integration sites using a PCR-based assay that has a much higher sensitivity and resolution than the agarose gel assay (Pryciak and Varmus, 1992).

10 *PCR assay.* One microgram of plasmid pBS-LA was incubated with the protein on ice for 5 min in the standard reaction buffer. The integration reaction was started by adding 15 nM of preprocessed U5 DNA (B2-1/V2) and incubating the sample at 37°C. After 30 or 60 min, the reaction was stopped by the addition of a final concentration of 15 mM EDTA. The sample was extracted with phenol-chloroform, ethanol precipitated in the presence of 10 µg tRNA, and washed with 70% ethanol. The pellet was resuspended in 50 µl of 10 mM Tris-HCl and 1 mM EDTA, pH 7.5. A 5 µl-aliquot of the reaction mixture was amplified for 25, 27, or 30 cycles of PCR: 1 min at 94°C, 1 min at 55°C and 2 min at 72°C. For analysis of the integration events occurring in the plus strand of the plasmid DNA, the PCR primers used were 0.2 µM unlabeled B2-1, 0.05 µM 5'-end labeled B2-1 and 0.25 µM BS+ (5'-CATTAAATGCAGCTGGCACGA-3', SEQ ID NO: 18), which is complementary to the plus strand of the plasmid DNA and is located at 232 bp from the 3'-end of the LexA-binding sequence. For analysis of the integration events occurring in the minus strand, the BS+ primer was replaced by the primer BS- (5'-TAATACGACTCACTATAGGG-3', SEQ ID NO: 19), which is complementary to the minus strand of the plasmid DNA and is located at 140 bp from the 3'-end of the LexA-binding sequence. The PCR reaction was performed in a buffer containing a final concentration of 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.001% w/v gelatin, 1.5 mM MgCl₂, 200 µM dNTPs, and 1 unit *Taq* polymerase (Perkin-Elmer Corp., Norwalk, CT), in a final volume of 20 µl. The labeled PCR products were analyzed on a denaturing 5% polyacrylamide gel and visualized by autoradiography.

15

20

25

30

Each band on the resulting autoradiogram corresponded to an integration event at a given phosphodiester bond. The frequency of integration at a particular site and its exact position was determined by the intensity of the band and by use of a sequencing ladder, respectively. Using the PCR assay, the distribution and frequency of integration events around the LexA-recognition sequence were compared between WT IN and IN1-288/LA. In the case of WT IN, with the LexA-binding site absent (pBS) or present (pBS-LA) in the target DNA, the distribution and intensity of the PCR-amplified products showed that most positions on the DNA could be used as target sites for integration, and there was a wide variation in integration frequency among the target sites.

With the fusion protein IN1-288/LA, when LexA-binding sequence was absent in the target DNA, the integration pattern was similar to that of the WT IN. When LexA-binding sequence was present in the target DNA, in contrast to the WT IN, the LexA-binding region was not used as a target by the fusion protein, and a majority of the integration events instead occurred near the regions flanking the LexA-binding sequence. Concurrently, there was a notable decrease in the frequency of integration in the outlying region (30 bp or more) of the LexA-binding sequence. Several integration hot spots located within 30 bp from the LexA-binding site, were found on the plus and minus strands of the target DNA. These hot spots were specific for the fusion protein and were not used as active target sites by the WT IN.

As a negative control, the integration reaction was carried out in the presence of a fixed amount of WT IN and various amounts of LexA protein. As the concentration of LexA protein increased in the reaction, there was a proportional decrease in the integration events occurring in the LexA-binding sequence. However, in contrast to the integration pattern observed with IN1-288/LA, there was no increase in integration in the regions flanking the LexA-binding sequence, nor a decrease in integration in the outlying regions. The data show that the integration pattern of IN1-288/LA results from two components working in *cis*, and not from a combined effect of two separate functions provided in *trans* by individual components.

Integration reaction using the PCR assay was also performed with the fusion protein IN1-288/LABD in order to examine possible differences in the integration pattern between fusion proteins containing full-length or only the DNA-binding domain of LexA protein. The integration pattern of IN1-288/LABD was similar to that of IN1-288/LA, except that the pattern of IN1-288/LABD was less specific since there was more integration within the LexA-binding sequence as well as the outlying regions. The result is consistent with the findings from the agarose gel assay and the footprinting analysis.

EXAMPLE 6

Target Site Usage of Truncated Integrase-LexA Fusion Proteins

The present example provides studies that examine whether truncated forms of integrase are competent at the integration function. The central core region of integrase contains the catalytic domain and the C-terminus of the protein is reported to bind non-specific DNA. To determine the minimal domain required for the preferred integration and to test whether higher specificity could be achieved by using an integrase without the non-specific DNA-binding domain, the integration patterns of fusion proteins containing various truncations of integrase by the PCR assay were examined.

The integration reaction was carried out for 1 h at 37°C in the presence of 250 nM of IN50-234, IN50-234/LA, IN50-288/LA, and IN1-234/LA. The recombinant products were amplified by PCR using oligonucleotides B2-1 and BS+ as primers. Twenty-seven cycles of PCR were performed for IN50-288/LA and IN1-234/LA, and 30 cycles for IN50-234 and IN50-234/LA. A control integration reaction was performed in the absence of protein, and subsequently amplified by 30 cycles of PCR.

The integration efficiencies of the truncated integrases, either by themselves or as fusion proteins, were approximately 100-fold lower than their full-length

counterparts. Other than the poor efficiency, the integration patterns of the truncated integrases IN50-234 and IN1-234 were unexpectedly similar to that of WT IN. Likewise, the integration patterns of fusion proteins containing a truncated integrase, such as IN50-234/LA, IN50-288/LA, and IN1-234/LA, were similar to that of IN1-288/LA. The close similarity of the integration patterns determined by the PCR-based assay between IN1-288/LA and the various truncated integrase-LexA fusion proteins indicate that no added specificity was achieved by removing the N- or C-terminus of integrase. The result indicates that though the C-terminus contributes to non-specific DNA binding, it is unlikely to be involved in target site selection. The result on the integration pattern of the truncated integrases suggests that the integrase domain responsible for target site selection may reside in the central core (amino acid residues from about 50-234, or about 50-212) of the protein.

EXAMPLE 7

D116N Integrase-DNA Binding Protein Domain Fusion Proteins

The present example provides for a fusion protein having an integrase domain with an aspartic acid residue, previously thought to be critical for catalysis, replaced with an asparagine residue. These studies demonstrate the utility of the present invention using a variety of substituted forms of the fusion protein.

The truncated integrases IN1-234 and IN50-234 showed a weak 3'-end joining activity when assayed by the sensitive PCR-based method; no 3'-end joining activity was detectable using the conventional *in vitro* assays. A weak 3'-end joining activity was also observed by the same PCR assay with a D116N mutant, which contains an asparagine substituting the highly conserved aspartic acid at position 116. The weak 3'-end joining activity observed with the truncated integrases and the D116N mutant was not changed in the presence or absence of the N-terminal His-tag. The D116N mutant has been shown previously to be inactive in all known catalytic activities of integrase using the conventional assays (Engelman and Craigie, 1992; Kulkosky *et al.*, 1992; Leavitt *et al.*, 1993; van Gent *et al.*, 1992).

Control experiments were carried out to confirm that the observed 3'-end joining activity of the truncated integrases and D116N mutant was not due to a contamination of the PCR. The similarity among the mutant and wild-type integrases in the banding pattern on a sequencing gel further supports that the PCR-amplified products were not experimental artifacts and that the truncated integrases and D116N mutant indeed possess 3'-end joining activity. This finding has important significance for *in vivo* experiments in which putatively integration-defective viruses are studied. In light of the weak 3'-end joining activity of the D116N mutant, it is possible that viruses containing a D116 mutation of integrase may be capable of forming a low level of proviruses, which may in turn produce sufficient Tat protein required for the indicator cell assay.

EXAMPLE 8

Feline Immunodeficiency Viral Integrase-DNA Binding Protein Domain Fusion Proteins

The present example provides a further fusion protein construct where the integrase catalytic domain is from feline immunodeficiency virus. The feline immunodeficiency virus (FIV) full-length integrase gene was obtained from plasmid p34TF10 (Talbot, *et al.*, 1989, provided by Tom Phillips at Scripps Research Institute) and was amplified by polymerase chain reaction (PCR). The 5' and 3' oligonucleotide primers for FIV integrase are 5'-CCAGTGCATATGTCCTCTTGGGTTGACAGA-3' and 5'-CAGTCAGGTACCCTCATCCCCTTCAGG-3' and contain *Nde* I and *Kpn* I sites at the N- and C-termini, respectively. After PCR, the DNA fragment containing the integrase gene was cut with *Nde* I and *Kpn* I. The cleaved DNA fragment was purified and ligated to pT7-7(His)/H-IN/LA plasmid DNA, previously cut with *Nde* I and *Bam*H I. The plasmid pT7-7(His) is derived from pT7-7, a T7 RNA polymerase-promoter system (Tabor and Richardson, 1985), and it contains an ATG initiation codon and seven histidine codons that are in-frame with the unique *Nde* I site. The

DNA sequence of the fusion construct was confirmed by dideoxy sequencing and the construct was transformed into *E. coli* BL21 (DE3).

5 The fusion protein was expressed under IPTG induction, and purified by nickel-chelating affinity chromatography and gel filtration chromatography. The purified FIV integrase-LexA fusion protein was catalytically active when tested by conventional *in vitro* assays (Vincent *et al.*, 1993; Chow and Brown, 1994); it was capable of carrying out 3'-end processing, 3'-end joining, and disintegration.

10 In addition to performing the functional assays, a PCR-based assay as described in Example 5 was utilized to determine if there was a bias in the selection of target sites towards the LexA DNA-binding sequence. The target substrate was a plasmid DNA containing a single binding site (LexA operator) for the LexA protein. The enzyme was first incubated with a preprocessed U5 viral DNA end to allow the integration reaction
15 to proceed. The reaction products were then subjected to PCR to determine at what locations integration had occurred. The PCR reaction was carried out with a radiolabeled primer to the U5 viral DNA substrate, and a primer approximately 250 bases downstream from the Lex A operator. In the presence of wild-type FIV IN, it was observed that integration occurred over a wide range of sites over the target DNA, with
20 no preferred integration site. However, integration of the viral DNA by the fusion protein exhibited a bias toward the DNA flanking the LexA operator. The directed integration mediated by the fusion protein required the presence of the LexA operator. This indicates that the LexA portion of the fusion protein is able to bind to the target sequence, and that integrase can then integrate into the adjacent DNA.

25 This construct would be particularly useful for human gene therapy protocols since the feline immunodeficiency virus is nonpathogenic for humans. In the construction of vector-host delivery systems where retroviruses are used as the vectors, there is some risk that the retrovirus may cause disease, and therefore, a nonpathogenic
30 feline virus construct would carry less risk of disease.

Another important reason for choosing FIV as the retroviral vector for site-directed integration is the availability of cats as an animal model for testing the feasibility of *in vivo* gene targeting in future studies.

5 *Preparation and catalytic activity of a truncated FIV integrase (I-235)/LexA fusion protein* -- In a separate study (Shibagaki, *et al.*, 1996), the C-terminal domain of FIV integrase (amino acid residues 236-281) was reported to be dispensable for its activity. A construct containing the truncated FIV integrase fused to LexA protein was prepared and tested to determine whether it possesses an increased specificity. The
10 truncated FIV integrase (I-235)/LexA gene was cloned into pT7-7 (His) using PCR amplification. The 5' primer for FIV INI-235 is identical to that described earlier for the full-length FIV integrase; the 3' primer is 5'-GCTAGAGGTACCTTTCTTATCTTTTGATC and contains a *Kpn* I site. After PCR, the DNA fragments containing the truncated integrase gene were cut with *Nde* I and
15 *Kpn* I. The cleaved DNA fragments were purified and ligated to pT7-7(His)/F-IN/LA plasmid DNA, previously cut with *Nde* I and *Kpn* I, and purified to remove the full length FIV integrase gene. The DNA sequence of the fusion construct was confirmed by dideoxy sequencing and the construct was transformed into *E. Coli* BL21 (DE3). The protein was expressed under IPTG induction, and purified by nickel-chelating
20 affinity chromatography and SP-sepharose chromatography.

The purified F-INI-235/LA fusion protein was catalytically active when tested by conventional *in vitro* assays; it was capable of carrying out 3'-end processing, 3'-end joining, and disintegration. Preliminary results obtained from the PCR-based assay
25 showed that integration of donor DNA mediated by the fusion protein containing a truncated FIV integrase, F-INI-235/LA, is also biased towards LexA-binding sequence. The relative specificity between the full-length and truncated fusion proteins is still under investigation. However, unlike the case with HIV-1 integrase, the activity of the F-INI-235/LA was only 2 to 3-fold less than that of the full-length integrase fusion
30 protein.

EXAMPLE 9
Integrase-DNA Binding Protein Domain
Fusion Proteins

5

The present example provides for a variety of DNA binding domains that may be fused to an integrase catalytic domain for purposes of the present invention.

10 In addition to *E. coli* LexA repressor protein and the reverse tetracycline repressor protein, several other sequence-specific DNA-binding proteins are suitable for forming a fusion protein with integrase. These further DNA-binding proteins and literature references in which sequences and/or plasmid sources may be found include (the references are incorporated by reference herein for this particular purpose): i) the tetracycline repressor of *E. coli* (Gossen and Bujard, 1992; Gossen *et al.*, 1995), ii) the
15 Lac repressor of *E. coli* (Reznikoff, 1992; Brown *et al.*, 1987), iii) GAL4 protein of yeast (*S. cerevisiae*) (Laughon and Gesteland, 1984), and iv) Cro repressor of phage lambda (Ohlendorf *et al.*, 1982; Hochschild and Ptashne, 1986).

20 These further DNA binding proteins or binding domains thereof will be fused to the C-terminus of integrase or to the C-terminus of an integrase catalytic domain in a similar manner to the strategy used for the integrase-LexA fusion protein as described in Example 1.

EXAMPLE 10
Expression Systems for Integrase-
25 **DNA Binding Protein Domain Fusion Proteins**

The present example provides expression vectors, and host cells for the expression of fusion proteins of the present invention.

30

To examine the generality of fusing integrase with other sequence-specific DNA-binding protein, a fusion protein consisting of full-length HIV-1 integrase and the reverse tetracycline repressor (rTET) of *E. coli* (Gossen, *et al.*, 1995) was prepared.

The N-terminus of rTet was fused to the C-terminus of HIV-1 integrase. The *rTet* gene was obtained by PCR amplification using pUHD172-Inco as the template. The 5' and 3' primers for the *rtet* gene are 5'-CAGTCAGGTACCTCTAGATTAGATAAAAGT-3' (SEQ ID NO:33) and 5'-CAGTCAGGATCCGGACCCACTTTTCACATTT-3', (SEQ ID NO: 34) respectively, and contain a *Kpn* I and *Bam*H I site, respectively. The PCR-amplified fragment was digested with *Kpn* I and *Bam*H I and cloned into pINI-288/LA previously cut with *Kpn* I and *Bam*H I. The fusion protein was purified according to the procedure described in Example 2, and the activities examined as described in Examples 2-5. The target DNA for IN/rTet fusion protein was pUHC13-3, which contains heptamerized Tet-operator sequences for specific binding of rTet. The result shows that integrase from different sources, such as HIV-1 and FIV, can be fused with different DNA-binding proteins, such as LexA and rTet, to achieve site-directed integration

Prokaryotic and eukaryotic cells useful for propagating vectors carrying a fusion protein gene of the present invention and for expression of the fusion protein include *E. coli* (e.g. BL21 (DE3), HB101, DH5 α), yeast such as *Pichia pastoris* (e.g. GS115) and *S. cerevisiae* (e.g. AB116), and insect cells (e.g. Sf9). The expression vectors useful for expression and purification of the fusion protein include pT7-7, pET, pBS24Ub, pYes2, and pAC360. Most preferably, the expression vector and the prokaryotic cell employed to propagate and express the fusion protein of the present invention are pT7-7 and *E. coli* BL21(DE3), respectively.

For ease of purification, the fusion protein of the present invention was purified with a histidine-tag (His-tag; sequence is a methionine followed by seven histidine residues) fused to the N-terminus of integrase. Inserted between the integrase and the His-tag was a thrombin cleavage site. Other peptides that can be fused to the N-terminus of integrase for the purpose of purification include glutathione-S-transferase, maltose-binding protein, and thioredoxin (Ausubel *et al.*, 1995). After purification, if necessary, the His-tag can be removed by thrombin digestion. The peptides for

purification can also be fused to the C-terminus of the LexA component of the fusion protein.

5 Fusion proteins will also be expressed in mammalian cell lines. Examples include VERO, HeLa cells, W138, COS, HOS, Jurkat, CEM, 293T and MDCK cell lines. Most preferably, a mammalian cell line employed to propagate an expression vector and for the expression of the fusion proteins of the present invention is 293T cells.

10 Expression vectors for mammalian cells useful for the expression of fusion proteins of the present invention include pCDM8, pZeoSV, pEUK-C1, pMAM, pREP, and pEBVHis. These vectors contain promoters (e.g. CMV, MMTV, RSV, SV40) for driving the expression of the cloned gene, polyA signal for termination of transcription, origin of replication (SV40, *oriP*), and selectable markers (e.g. resistance to neomycin, 15 hygromycin, and zeocin).

EXAMPLE 11
Targeted Delivery of Integrase-
20 **DNA Binding Protein Domain Fusion Proteins**

The present example provides for targeted delivery of a fusion protein of the present invention.

25

For site-directed integration of a donor DNA using a fusion protein that contains a C-terminal LexA binding domain, the nucleotide sequence representing the LexA binding site may be introduced into the target DNA. This allows the use of the fusion protein having a LexA binding domain for the integration of virtually any donor DNA 30 into any target DNA. In particular, these reagents may be supplied as laboratory reagents for that purpose. The LexA binding site is most easily introduced into a target DNA at a restriction enzyme site, where the appropriate linkers have been attached to

the ends of the double stranded LexA binding site oligonucleotide molecule. The LexA-binding site may also be introduced by homologous recombination (Bollag *et al.*, 1989). In such an approach, the LexA-binding sequence will be flanked by DNA sequences homologous to the region of insertion.

5

Using similar methods, any nucleotide sequence that represents a binding site on DNA may be introduced into a target DNA, and the corresponding DNA binding domain having binding specificity for that DNA sequence is engineered into a fusion protein.

10

There are numerous ways for introducing a donor DNA and the fusion protein into target cells (cells that receive targeted integration) including electroporation, microinjection, calcium phosphate coprecipitation, liposome-based membrane fusion, and use of adenoviral vectors. In the present invention, the preferred means is via retroviral vectors. The first step of the process is to produce infectious, yet replication-defective viruses. There are two general methods for doing so. In the first method, a stable helper cell line will be prepared by transforming 293T cells with a plasmid containing a partial retrovirus genome. The partial genome contains the essential genes, *gag*, *pol*, *env*; and the integrase gene at the 3' end of the *pol* gene is substituted with a gene encoding a fusion protein of the present invention. The partial viral genome lacks the packaging signal and the psi sequence, so the RNA transcribed from the viral genome cannot be packaged into viral particles. The function of the helper cell is, therefore, to provide essential viral proteins and the fusion protein so that a donor DNA of choice can be packaged. To this helper cell, a donor retroviral DNA vector will be introduced. Commonly used retroviral vectors include LNSX, LNCX, LHDCX, LXSHD, and LXSH (Miller *et al.*, 1993). Many of these vectors contain DNA sequences derived from murine leukemia virus (MLV). Essentially, the donor vector DNA contains the LTR (which contains the sequences for integration), the packaging signal, a selectable marker (e.g. neomycin resistance), and a promoter upstream of a site for gene insertion. The gene inserted can be any gene of interest, for example, the adenosine deaminase gene. For safety reasons, the retroviral vector does not contain

15

20

25

30

any essential viral genes. The necessary viral proteins deleted from the disabled vector must be therefore provided "in *trans*" by the helper cell. Since the RNA transcribed from the retroviral vector has the packaging signal, it will be packaged by the viral proteins provided by the helper cell to form infectious, replication-defective viruses, which can be harvested from the culture medium.

Many cell lines, known to one of skill in this art in light of this disclosure, contain viral functions necessary for packaging and delivery of replication-defective viral vectors derived from several commonly used tumor viruses. These useful viruses include MLV, spleen necrosis virus (SNV), avian leukosis virus (ALV), and reticuloendotheliosis virus (REV). Patents have issued for helper cell lines for MLV and REV (Miller, U.S. Pat. No. 4,861,719; Temin *et al.*, U.S. Pat. No. 4,650,764). These existing helper cell lines, of course, do not contain a gene that encodes a fusion protein of the present invention, however, they can be modified to carry a fusion protein-encoding gene.

MLV viruses have become useful vectors for animal genetic engineering of cells and organisms, because of their compatibility with a wide variety of animal cell types including certain germ cells as well as human cells. MLV was used to insert viral transgenes into the mouse germline, creating a transgenic mouse (Jaenisch *et al.*, 1976, 1981). MLV vector systems have been approved for limited human gene therapy trials despite some of the problems described previously.

In a further method, a helper cell is not prepared. Instead, the plasmid DNA containing the essential viral genes and the plasmid containing the donor retroviral vector will be co-transfected into 293T cells. The replication-defective viruses will then be harvested from the culture medium. In both methods, the replication-defective retroviruses, which contain the donor RNA and the fusion protein, will be used to infect target cells.

It is envisioned that the replication-defective virus, prepared by the methods described earlier, will be used to introduce a donor RNA containing a therapeutic gene into a host cell. After infection, the donor RNA will be made into cDNA by the viral reverse transcriptase. The donor cDNA will then enter the nucleus and integrate into a specific site determined by the specificity of the DNA-binding moiety of the fusion protein.

A modified FIV containing the integrase/LexA fusion will be prepared to produce infectious, replication-defective retroviruses for site-directed integration as an *in vivo* representative model. The approach involves the use of a replication-defective virus, FIV Δ E-N, which is derived from the full-length FIV clone or f2rep (Scripps Research Institute). FIV Δ E-N contains a deletion (map positions 7248-8287) in the *env* gene, and the deleted fragment will be replaced with a neomycin-resistant gene. The plasmid DNA containing the FIV Δ E-N will be digested with *Bsp* H I and *Avr* II, which cleave the genome within the integrase gene at positions 4436 and 6718, respectively. The FIV integrase/LexA fusion gene will be amplified by PCR, and the product partially digested with *Bsp* H I and *Avr* II. The desired fragment will be isolated and ligated with the similarly cleaved FIV Δ E-N to produce FIV Δ N Δ E-N. The final construct retains all the known splice donor and acceptor sites, and the putative *vif* and *rev* genes of FIV that are required for gene expression and infectivity (Talbot, *et al.*, 1989). The replication-defective virus will be pseudotyped with the envelope of vesicular stomatitis virus. A virus stock will be generated by electroporation of 293T cells at 50% confluence using 10 μ g of FIV Δ N Δ E-N plasmid DNA and 10 μ g of envelope-expressing plasmid DNA. The culture supernatant will be collected and filtered 60 h later. The virus stock will be titered and characterized by measuring the p25 (capsid) content and the *in vitro* reverse transcriptase activity. The ability of the fusion protein to mediate site-directed integration in tissue culture cells will be examined by using the pseudotyped, modified FIV (FIV Δ N Δ E-N) to infect HeLa cells that have previously been infected with SV40. The SV40 used contains a wild-type or mutated LexA operator site inserted into the unique *Kpn* I site located in the noncoding region of the 5.2 kbp genome. SV40 DNA was chosen as a target because SV40

replicates to a copy number of about 10^5 , which makes it possible to analyze many thousands of integration events from a single experiment. The use of extrachromosomal DNA as a target will also lower the nonspecific amplification that can result from using the genomic DNA. The recombinant products will be separated
5 from the chromosomal DNA, and the distribution of the integration sites used *in vivo* will be determined by the assays described earlier in Examples 2-5.

U.S. Patent 5,399,346 to Anderson *et al.* is incorporated by reference herein as teaching gene therapy techniques, particularly methods whereby primary human cells
10 are genetically engineered with DNA (RNA) encoding a therapeutic which is to be expressed *in vivo*.

EXAMPLE 12

15 **Integrase Fusion Proteins where the N-terminal
Zinc Finger Domain is Substituted by a DNA Binding Domain**

The present example provides another potential approach for engineering
20 integration proteins having site-specificity for binding to DNA. The present inventors envision the replacement of the N-terminal zinc-finger motif of integrase (from about amino acids 1-50) with other zinc-finger protein domains having binding specificity for DNA sequences (Berg, 1990; Klug and Rhodes, 1987). In this approach, the zinc-finger motif of integrase will be deleted and replaced with other zinc-finger motif that
25 recognizes specific DNA sequences. By exchanging the zinc-finger motif, the resulting hybrid protein may retain the integration activity and may gain an added ability to recognize specific DNA sequences.

EXAMPLE 13

30 **Further Integrase Constructs**

The integrase-LexA fusion protein of the present invention has binding specificity for an *E. coli* LexA nucleotide sequence and would not be normally expected

to bind specifically to a human DNA sequence. However, considering the size of the human genome of 3 billion bp, the integrase-LexA protein may bind to several LexA-like sequences in the genome. Integration into these LexA-like sequences may be harmless, alternatively, the LexA-binding sequence may be introduced into a desired target site for specific integration.

The present example addresses this aspect and provides for further integrase constructs, for example, a construct where an N-terminal integrase catalytic domain is fused to a protein domain having affinity for a transcription factor, and a construct where an integrase is covalently bonded to an oligonucleotide which provides binding specificity for its complementary nucleotide sequence.

Integrase Fused to RNA Polymerase III Transcription Factor -- RNA polymerase III (Pol III) is responsible for transcribing tRNA and some small nuclear RNA genes. Transcription by Pol III involves the polymerase itself and several protein factors called transcription factors, such as TFIIIA, TFIIIB, and TFIIIC. TFIIIB is believed to be recruited to the transcription complex by its interaction with TFIIIC and Pol III. TFIIIB itself is a large complex and contains many subunits. One subunit is BRF (IIIB-related factor). The present inventor envisions a fusion protein consisting of integrase and BRF. In such a strategy, the fusion protein will be brought into close proximity of Pol III transcribed genes through protein-protein interaction (BRF and TFIIIC and Pol III). Advantages of such an approach are i) protein-protein interaction may be more specific than protein-DNA interaction, ii) integration would likely be directed towards regions that are transcribed by Pol III, which most likely are tRNA genes. These regions are ideal sites because i) they are transcriptionally active, and ii) tRNA genes are in multiple copies, and disruption of one tRNA gene by integration should not have a detrimental effect on the cell.

Integrase Covalently Linked with an Oligonucleotide -- In this approach, an oligonucleotide will be covalently linked to an amino acid residue of integrase, possibly through an amide bond with aspartic acid or glutamic acid, or a disulfide linkage with

a cysteine. Site-directed integration will be achieved by base-pairing between the oligonucleotide of the integrase-linked oligonucleotide and the complementary region of the genome. The main advantage of this strategy is that any region of the genome can be targeted as long as some information on the DNA sequence of the desired region is known. This approach is particularly applicable to *ex vivo* gene therapy.

EXAMPLE 14
Purging of Stem and Cord Blood Cells with
Fusion Protein Mediated Gene Transfer

The present example provides a description of potential uses of the herein described site-specific integration of DNA into stem or cord blood cells *ex vivo*. Stem cells are obtained from a patient in need of gene therapy, for example, a patient having cancer, particularly leukemia, AIDS, or a genetic disease. Cord blood cells are obtained from placenta. Stem cells or cord blood cells are treated with a replication-defective retrovirus harvested from helper cells encoding a fusion protein of the present invention and with donor DNA. Treated stem or cord blood cells are transferred to the patient to provide a transplant.

Donor DNA in this case may be genes for therapeutic replacement of defective genes, genes for providing a therapeutic function, or DNA for disruption of an undesirable gene. Examples include providing a gene encoding clotting factor VIII or IX for hemophilia, the *ada* gene for adenosine deaminase deficiency, a gene encoding the chloride channel for cystic fibrosis, or an LDL receptor encoding gene for hypercholesterolemia.

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may

be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described
5 herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated
5 herein by reference.

- Ausubel, F.M. *et al.*, (eds), *Current Protocols in Molecular Biology*, 1995, John Wiley & Sons, New York
- Berg, J., *J. Biol. Chem.* 265: 6513-6516, 1990.
- 10 Bollag, *et al.*, 1989, *Ann. Rev. Genet.* 23:199-225.
- Brenowitz, M., *et al.*, 1993, "Footprinting of nucleic acid-protein complexes, p. 1-43. In A. Revzin (ed.), *Quantitative DNase I Footprinting*. Academic Press, Inc. (p. 1-43).
- Brenowitz, M., *et al.*, 1986, *Methods Enzymol.* 130:132-181.
- 15 Brent, R. and M. Ptashne, 1985, *Cell* 43:729-736.
- Brent, R. and M. Ptashne, 1981, *Proc. Natl. Acad. Sci. USA* 78:4204-4208.
- Brown, M., *et al.*, *Cell* 49: 603-612, 1987.
- Brown, P.O., 1990, *Microbiol. Immunol.* 157:19-48.
- Bushman, F.D., 1994, *Proc. Natl. Acad. Sci. USA* 91:9233-9237.
- 20 Bushman, F.D. and B. Wang, 1994, *J. Virol.* 68:2215-2223.
- Bushman, F.D., *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:3428-3432.
- Cannon, P., *et al.*, *J. Virol.*, 1994, 68:4768-4775.
- Caracciolo *et al.* (1989) *Science*, 245:1107.
- Chow, S. and P. Brown, *J. Virol.* 68: 3896-3907, 1994.
- 25 Chow, S. A., *et al.*, 1992, *Science* 255:723-726.
- Craigie, R., 1992, *Trends Genet.* 8:187-190.
- Dumoulin, P., *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:2030-2034.
- Engelman, A. and R. Craigie, 1992, *J. Virol.* 66:6363-6369.
- Engelman, A., *et al.*, 1994, *J. Virol.* 68:5911-5917.
- 30 Fitzgerald, M.L. and D.P. Grandgenett, 1994, *J. Virol.* 68:4314-4321.
- Fogh, R.H., *et al.*, 1994, *EMBO J.* 13:3936-3944.

- Goff, S.P., 1992, *Annu. Rev. Genet.* 26:527-544.
- Golemis, E.A. and R. Brent, 1992, *Mol. Cell. Biol.* 12:3006-3014.
- Gossen, M., and H. Bujard, *Proc. Natl. Acad. Sci. USA* 89: 5547-5551, 1992.
- Gossen, M., *et al.*, *Science* 268: 1766-1769, 1995.
- 5 Grandgenett, D.P., *et al.*, 1993, *J. Virol.* 67:2628-2636.
- Hochschild, A., and M. Ptashne, *Cell* 44: 925-933, 1986.
- Jaenisch *et al.*, *Proc. Nat. Acad. Sci. (USA)* 73:1260, 1976.
- Jaenisch *et al.*, *Cell*, 24:519, 1981.
- Johnson, M.S., *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:7648-7652.
- 10 Kalpana, G.V., *et al.*, 1994, *Science* 266:2002-2006.
- Katzman, M. and M. Sudol, *J. Virol.*, 1994, 68:3558-3569.
- Kim, B. and J.W. Little, 1992, *Science* 255:203-206.
- Kitamura, Y., *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89:5532-5536.
- Klug, A., and D. Rhodes, *Trends Biochem. Sci.* 12: 464-469, 1987.
- 15 Kulkosky, J., *et al.*, 1992, *Mol. Cell. Biol.* 12:2331-2338.
- Laughon, A., and R.F. Gesteland, *Mol. Cell. Biol.* 4: 260-267, 1984.
- Leavitt, A.D., *et al.*, 1993, *J. Biol. Chem.* 268:2113-2119.
- Lewis, L.K., *et al.*, 1994, *J. Mol. Biol.* 241:507-523.
- Little, J.W. and D.W. Mount, 1982, *Cell* 29:11-22.
- 20 Little, J.W., *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:4199-4203.
- Merrifield, R., *J. Am. Chem. Soc.*, 85:2149, 1963.
- Miller, A.D., *et al.*, *Methods Enzymol.* 217:581-599, 1993.
- Muller, H.P. and H.E. Varmus, 1994, *EMBO J.* 13:4704-4714.
- Mulligan, R.C., 1993, *Science* 260:926-932.
- 25 Ohlendorf, D., *et al.*, *Nature* 298:718-723, 1982.
- Pruss, D., *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* 91:5913-5917.
- Pruss, D., *et al.*, 1994, *J. Biol. Chem.* 269:25031-25041.
- Pryciak, P.M., *et al.*, 1992, *EMBO J.* 11:291-303.
- Pryciak, P.M. and H.E. Varmus, 1992, *Cell* 69:769-780.
- 30 Ptashne, M, 1992, *In A Genetic Switch*. Cell Press and Blackwell, Cambridge, MA.

- Remington: The Science and Practice of Pharmacy, 19th edition, Volumes 1 and 2, A.R. Gennaro, ed., Mack Publishing Co. Easton, PA, 1995.
- Reznikoff, W., *Mol. Microbiol.* 6: 2419-2422, 1992.
- Rohdewohld, H., *et al.*, 1987, *J. Virol.* 61:336-343.
- 5 Sambrook, J., *et al.*, 1989, *Molecular Cloning: a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sandmeyer, S.B., *et al.*, 1990, *Annu. Rev. Genet.* 24:491-518.
- Schauer, M. and A. Billich, 1992, *Biochem. Biophys. Res. Comm.* 185:874-880.
- Schmidt-Dorr, T., *et al.*, 1991, *Biochemistry* 30:9657-9664.
- 10 Schnarr, M., *et al.*, 1988, *FEBS Letters* 234:56-60.
- Schnarr, M., *et al.*, 1991, *Biochimie* 73:423-431.
- Shibagaki, Y., *et al.*, 1996, *Virology* submitted
- Shiramizu, B., *et al.*, 1994, *Cancer Res.* 54:2069-2072.
- Tabor, S. and C.C. Richardson, 1985, *Proc. Natl. Acad. Sci. USA* 82:1074-1078.
- 15 Talbott, R., *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:5743-5747.
- Temin, H.M., 1990, *Hum. Gene Ther.* 1:111-123.
- Thliveris, A.T. and D.W. Mount, 1992, *Proc. Natl. Acad. Sci. USA* 89:4500-4504.
- van Gent, D.C., *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89:9598-9602.
- Varmus, H.E., and P.O. Brown, 1989, "Retroviruses", p. 53-108. In M. Howe and D.
- 20 Berg (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Vijaya, S., *et al.*, 1986, *J. Virol.* 60:683-692.
- Vincent, K. A., *et al.*, 1993, *J. Virol.* 67:425-437.
- Vink, C., *et al.*, 1993, *Nucleic Acids Res.* 21:1419-1425.
- Vink, C. and R.H.A. Plasterk, 1993, *Trends Genet.* 9:433-437.
- 25 Vojtek, A.B., *et al.*, (1993) 74: 205-214.
- Wang, H. and D.J. Stillman, 1993, *Mol. Cell. Biol.* 13:1805-1814.
- Wertman, K.F. and D.W. Mount, 1985, *J. Bacteriol.* 163:376-384.
- Withers-Ward, E.S., *et al.*, 1994, *Genes Dev.* 8:1473-1487.
- Woernier, A.M., *et al.*, 1992, *AIDS Res. Hum. Retroviruses* 8:2433-2437.

WHAT IS CLAIMED IS:

1. A fusion protein comprising a retroviral integrase catalytic domain COOH-terminally coupled to a DNA binding protein domain having binding specificity for a target nucleotide sequence, the fusion protein capable of integrating a donor DNA molecule into a target DNA molecule at or near the target nucleotide sequence.
2. The fusion protein of claim 1 wherein the retroviral integrase catalytic domain is integrase from human immunodeficiency virus type 1 or type 2.
3. The fusion protein of claim 1 wherein the retroviral integrase catalytic domain is from human immunodeficiency virus type 1 integrase.
4. The fusion protein of claim 1 wherein the retroviral integrase catalytic domain includes a sequence of amino acids from about amino acid 50 to about amino acid 212 of human immunodeficiency virus type 1 integrase.
5. The fusion protein of claim 1 wherein the retroviral integrase catalytic domain is from feline immunodeficiency virus integrase.
6. The fusion protein of claim 1 wherein the DNA binding protein domain having binding specificity for a target nucleotide sequence is from *E. coli* LexA repressor protein, reversed wild-type tetracycline repressor protein of *E. coli*, Lac repressor of *E. coli*, GAL4 protein of yeast, or Cro repressor of phage lambda.

7. The fusion protein of claim 1 wherein the DNA binding protein domain having binding specificity for a target nucleotide sequence is LexA binding protein domain.
- 5 8. The fusion protein of claim 7 where the target nucleotide sequence is CTGTNNNNNNNNACAG (SEQ ID NO:20).
- 10 9. The fusion protein of claim 1 having an amino acid sequence essentially as set forth in SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:29, SEQ ID NO:31, a combination thereof, or a biologically functional fragment thereof.
- 15 10. A purified nucleic acid molecule consisting essentially of a nucleotide sequence encoding the fusion protein of claim 1.
- 20 11. The purified nucleic acid molecule of claim 10 wherein the molecule is a DNA molecule and the nucleotide sequence is essentially as set forth in SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:28, SEQ ID NO:30, a combination thereof, or a biologically functional fragment thereof.
- 25 12. A vector comprising a nucleotide sequence encoding the fusion protein of claim 1.
- 30 13. The vector of claim 12 defined further as an expression vector having a promoter operatively linked to the nucleotide sequence.

14. The vector of claim 13 wherein the expression vector is pT7-7, pET, pBS24Ub, pYes2, or pAC360.

5 15. A host cell transformed to include a nucleotide sequence encoding the fusion protein of claim 1.

10 16. The host cell of claim 15 wherein the cell is a eukaryotic cell.

17. A method of integrating a donor DNA molecule at or near a specific site on a target DNA molecule comprising:

15 selecting a DNA binding protein domain having binding affinity for the specific site on the target DNA molecule;

constructing a fusion protein having an N-terminal retroviral integrase catalytic domain and the DNA binding protein domain at a C-terminus; and

20 contacting the donor DNA molecule, the target DNA molecule and the fusion protein,

wherein the fusion protein directs integration of the donor DNA molecule at or near the specific site of the target DNA molecule.

25 18. The method of claim 17 wherein the donor DNA molecule comprises a gene encoding an integrase-DNA binding moiety fusion protein.

30

19. The method of claim 17 wherein the donor DNA molecule comprises a gene encoding an integrase-DNA binding moiety fusion protein.
- 5 20. The method of claim 17 where the fusion protein has an amino acid sequence as defined in SEQ ID NO:23.
- 10 21. The method of claim 17 where the fusion protein has an amino acid sequence as defined in SEQ ID NO:25.
- 15 22. The method of claim 17 wherein the contacting step comprises the steps of:
incubating the fusion protein with the target DNA molecule to form an incubate; and
contacting the incubate with the donor DNA molecule.
- 20 23. The method of claim 17 wherein the target DNA is DNA containing a defective gene or DNA containing an oncogene.
- 25 24. The method of claim 17 wherein the retroviral integrase catalytic domain is integrase from human immunodeficiency virus type 1 or type 2, or feline immunodeficiency virus.

25. The method of claim 17 wherein the DNA binding domain protein is the LexA binding protein, and the specific site on the target DNA molecule is the LexA binding sequence.

5

26. A method of integrating a donor DNA molecule at or near a selected site on a target DNA molecule comprising

10

introducing a LexA nucleotide sequence at the selected site on the target DNA molecule to form a LexA target DNA molecule; and

contacting the donor DNA molecule, the LexA target DNA molecule and a fusion protein having an N-terminal retroviral integrase catalytic domain and a C-terminal LexA binding domain;

15

wherein the fusion protein facilitates integration of the donor DNA molecule into the target DNA molecule near the LexA target site.

20

27. The method of claim 26 where the LexA nucleotide sequence is CTGTATGAGCATACAG, (SEQ ID NO:21).

25

28. A method of inactivating an oncogene by integrating a donor DNA molecule at or near the oncogene, or regulatory regions thereof, comprising:

selecting a DNA binding protein domain having binding affinity for the oncogene or regulatory regions thereof;

30

constructing a fusion protein having an N-terminal retroviral integrase catalytic domain and the DNA binding protein domain at a C-terminus; and

contacting a donor DNA molecule, the oncogene or regulatory regions thereof,
and the fusion protein,

wherein the fusion protein facilitates integration of the donor DNA molecule at or near
5 the oncogene or regulatory regions thereof, thereby inactivating the oncogene.

29. A fusion protein comprising a catalytic domain of retroviral integrase and an N-
terminal zinc finger domain having binding specificity for a DNA molecule where the
10 zinc finger domain is other than a zinc finger domain naturally occurring with the
catalytic domain in a retroviral integrase molecule.

30. A fusion protein comprising an integrase catalytic domain fused to a protein
15 domain having affinity for a transcription factor.

31. A protein-oligonucleotide construct comprising an integrase catalytic domain
bonded to an oligonucleotide.

20

1/13

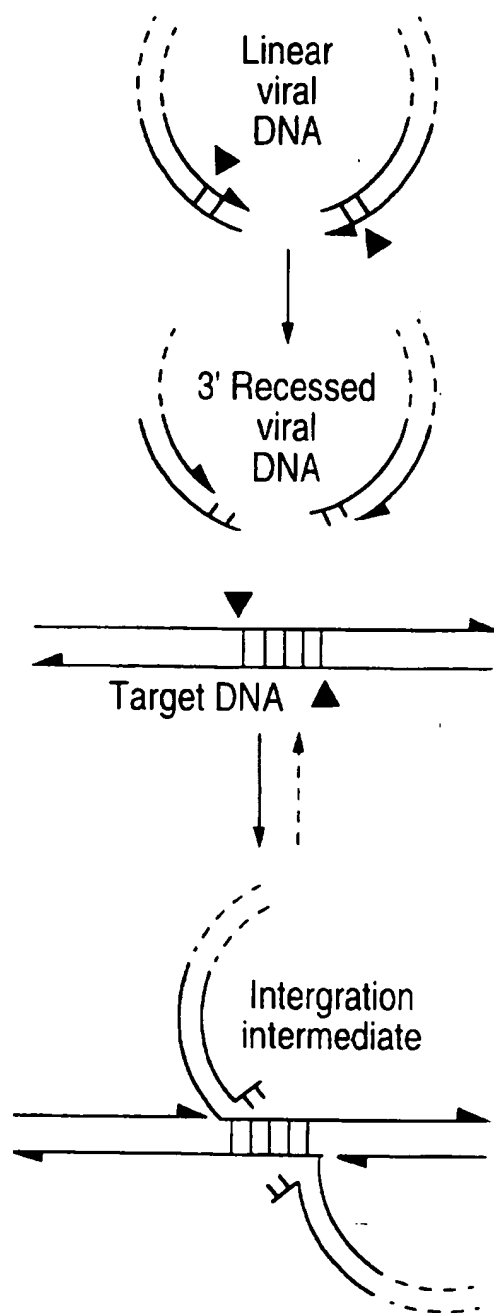


Fig. 1

2/13

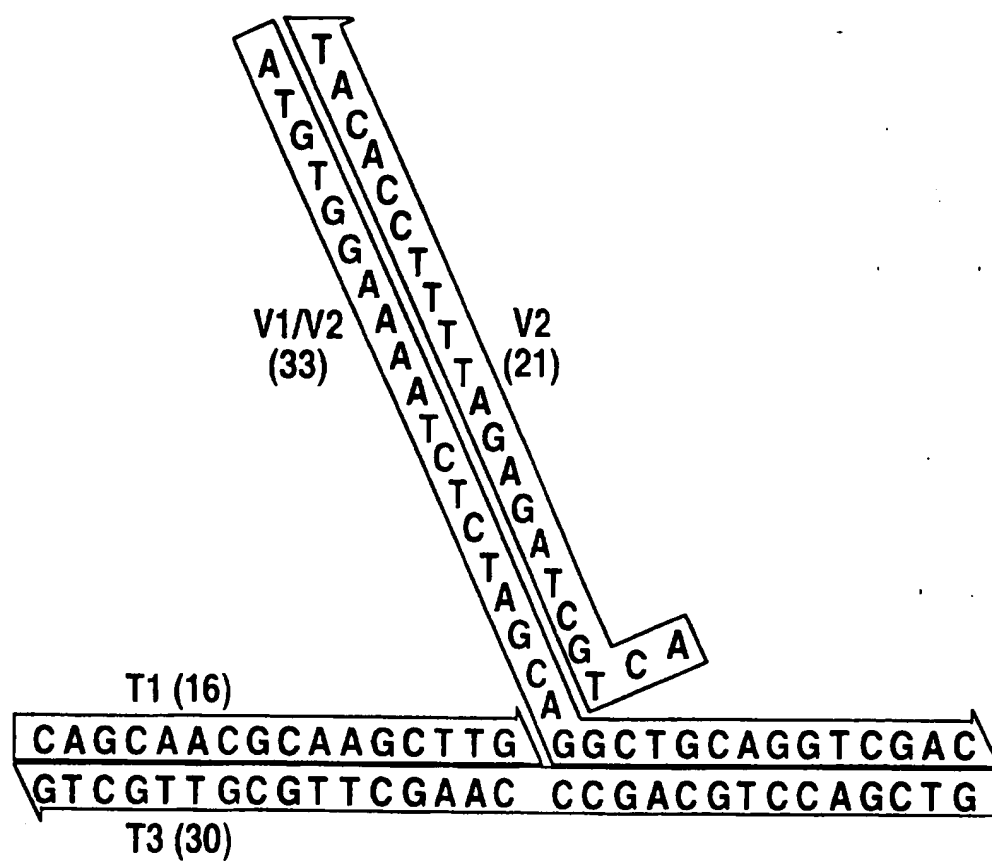


Fig. 2

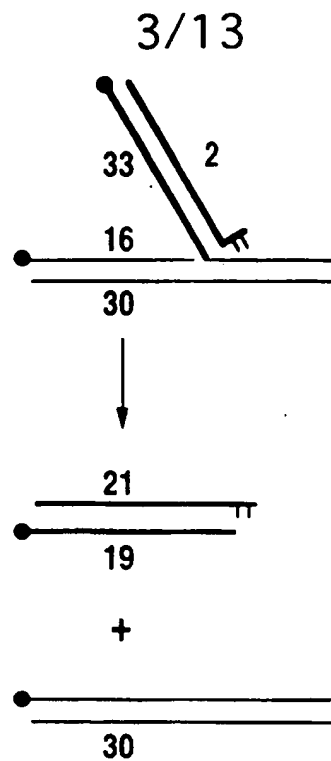


Fig. 3

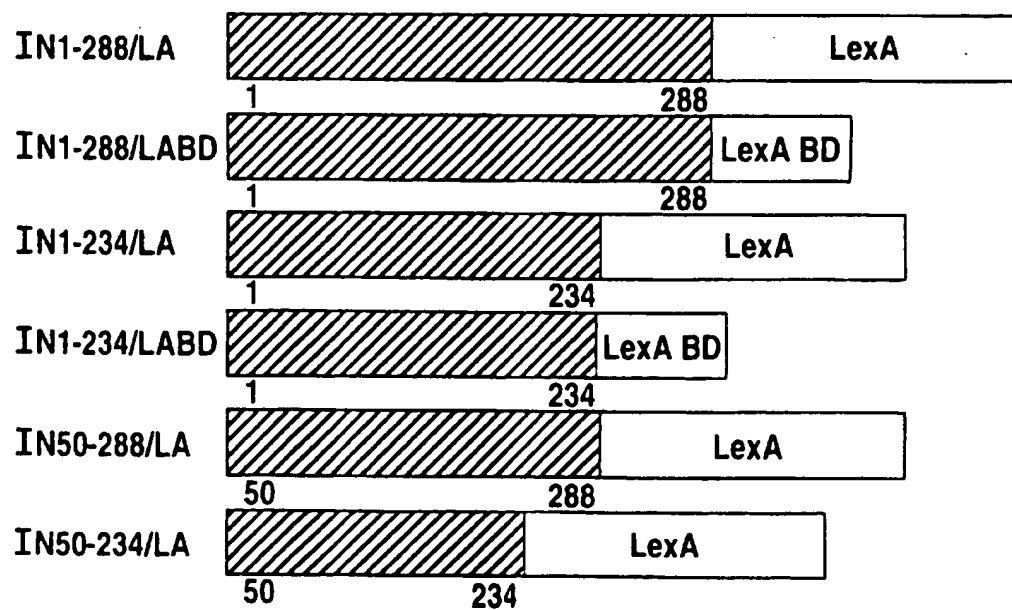


Fig. 4

4/13

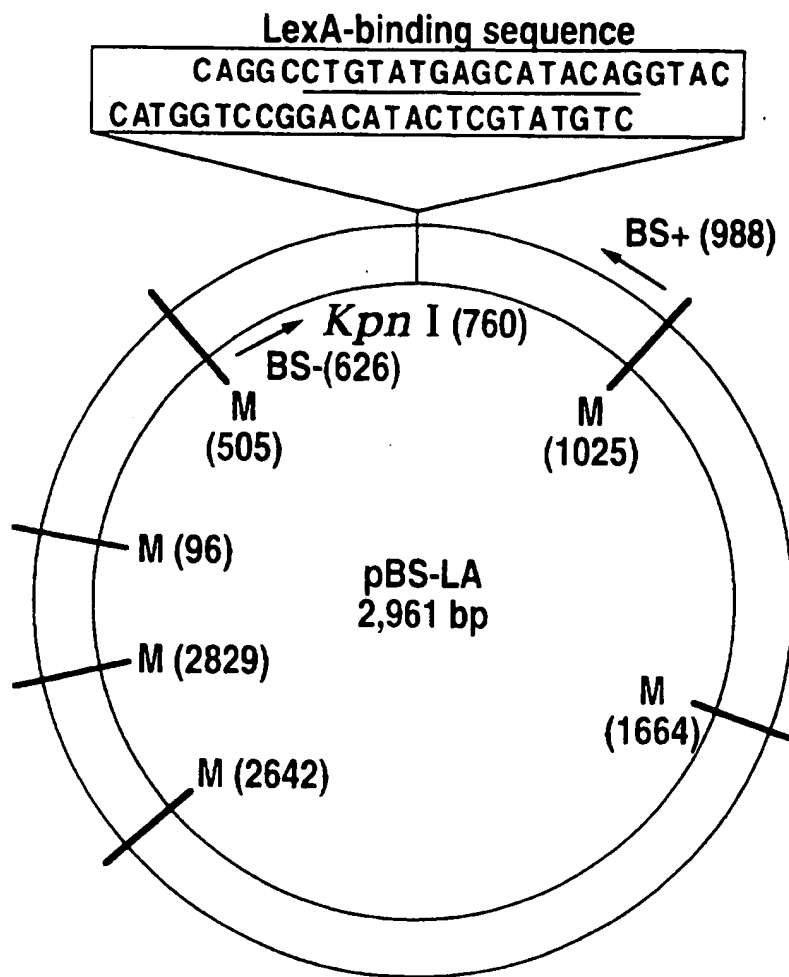


Fig. 5

5/13

ATGCATGGACAAGTAGACTGTAGCCCAGGAATATGGCAGCTAGATTGTACACATTTAGAA	60
M H G Q V D C S P G I W Q L D C T H L E	20
GGAAAAGTTATCTTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTAATT	120
G K V I L V A V H V A S G Y I E A E V I	40
CCAGCAGAGACAGGGCAAGAAACAGCATACTTCCTCTTAAAATTAGCAGGAAGATGGCCA	180
P A E T G Q E T A Y F L L K L A G R W P	60
GTA AAAACAGTACATACAGACAATGGCAGCAATTTACCAGTACTACAGTTAAGGCCGCC	240
V K T V H T D N G S N P T S T T V K A A	80
TGTTGGTGGGCGGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAGGA	300
C W W A G I K Q E F G I P Y N P Q S Q G	100
GTAATAGAATCTATGAATAAAGAATTAAGAAAATTATAGGACAGGTAAGAGATCAGGCT	360
V I E S M N K E L K K I I G Q V R D Q A	120
GAACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGG	420
E H L K T A V Q M A V F I H N F K R K G	140
GGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAA	480
G I G G Y S A G E R I V C I I A T D I Q	160
ACTAAAGAAGGTACCAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGATCTCATCCGT	540
T K E G T K A L T A R Q Q E V F D L I R	180
GATCACATCAGCCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAGCGTTTGGGG	600
D H I S Q T G M P P T R A E I A Q R L G	200
TTCCGTTCCCCAAACGCGGCTGAAGAACATCTGAAGGCGCTGGCACGCAAAGGCGTTATT	660
F R S P N A A E E H L K A L A R K G V I	260
GAAATTGTTTCCGGCGCATCACGCGGGATTCTGTCTGTTGCAGGAAGAGGAAGAAGGGTTG	720
E I V S G A S R G I R L L Q E E E E G L	280
CCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCA	753
P L V G R V A A G E P	291

Fig. 6

6/13

TTTTTAGATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGGAGA 60
F L D G I D K A Q E E H E K Y H S N W R 20

GCAATGGCTAGTGATTTTAACTACCACTGTAGTAGCAAAAGAAATAGTAGCCAGCTGT 120
A M A S D F N L P P V V A K E I V A S C 40

GATAAATGTCAGCTAAAAGGGGAAGCCATGCATGGACAAGTAGACTGTAGCCCAGGAATA 180
D K C Q L K G E A M H G Q V D C S P G I 60

TGGCAGCTAGATTGTACACATTTAGAAGGAAAAGTTATCTTGGTAGCAGTTCATGTAGCC 240
W Q L D C T H L E G K V I L V A V H V A 80

AGTGGATATATAGAAGCAGAAGTAATTCCAGCAGAGACAGGGCAAGAAACAGCATACTTA 300
S G Y I E A E V I P A E T G Q E T A Y F 100

CTCTTAAATTAGCAGGAAGATGGCCAGTAAAAACAGTACATACAGACAATGGCAGCAAT 360
L L K L A G R W P V K T V H T D N G S N 120

TTCACCAGTACTACAGTTAAGGCCGCCTGTTGGTGGGCGGGGATCAAGCAGGAATTTGGC 420
F T S T T V K A A C W W A G I K Q E F G 140

ATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAATAAAGAATTAAAGAAA 480
I P Y N P Q S Q G V I E S M N K E L K K 160

ATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCAGTA 540
I I G Q V R D Q A E H L K T A V Q M A V 180

TTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATA 600
F I H N F K R K G G I G G Y S A G E R I 200

GTAGACATAATAGCAACAGACATACAACTAAAGAATTACAAAAACAAATTACAAAAATT 660
V D I I A T D I Q T K E L Q K Q I T K I 260

CAAAATTTTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGAAAGGACCAGCAAAG 720
Q N F R V Y Y R D S R D P V W K G P A L 240

Fig. 7A

7/13

CTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTG 780
L L W K G E G A V V I Q D N S D I K V V 260

CCAAGAAGAAAAGCAAAGATCATCAGGGATTATGGAACAGATGGCAGGTGATGATTGT 840
P R R K A K I I R D Y G K Q M A G D D C 280

GTGGCAAGTAGACAGGATGAGGATGGTCCCAAAGCGTTAACGGCCAGGCAACAAGAGGTG 900
V A S R Q D E D G T K A L T A R Q Q E V 300
▲ ▲

TTTGATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGTGCGGAAATC 960
F D L I R D H I S Q T G M P P T R A E I 320

GCGCAGCGTTTGGGGTTCCGTTCCCAAACGCGGCTGAAGAACATCTGAAGGCGCTGGCA 1020
A Q R L G F R S P N A A E E H L K A L A 340

CGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTCTGTCTGTTGCAGGAA 1080
R K G V I E I V S G A S R G I R L L Q E 360

GAGGAAGAAGGGTTGCCGCTGGTAGGTCTGTGGCTGCCGGTGAACCACTTCTGGCGCAA 1140
E E E G L P L V G R V A A G E P L L A Q 380

CAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTC 1200
Q H I E G H Y Q V D P S L F K P N A D F 400

CTGCTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGTGACTTGCTG 1260
L L R V S G M S M K D I G I M D G D L L 420

GCAGTGCATAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTGCGACGTA TTGATGAC 1320
A V H K T Q D V R N G Q V V V A R I D D 440

CAAGTTACCGTTAAGCGCCTGAAAAACAGGGCAATAAAGTCGAACTGTTGCCAGAAAAT 1380
E V T V K R L K K Q G N K V E L L P E N 460

AGCGAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTG 1440
S E F K P I V V D L R Q Q S F T I E G L 480

GCGGTTGGGGTTATTCGCAACGGCGACTGGCTG 1473
A V G V I R N G D W L 491

Fig. 7B

8/13

ATTTTGTTTAACTTTAAGAAGGAGATATACATAATGCATCACCATCACCATCACCATATC
1 -----+-----+-----+-----+-----+-----+-----+ 60
TAAACAAATTGAAATTCTTCCTCTATATGTATTACGTAGTGGTAGTGGTAGTGGTATAG
M H H H H H H H I 9
His-Tag

GTTCCGCGTGGATCTATGTCCTCTTGGGTTGACAGAATTGAGGAAGCAGAAATAAATCAT
61 -----+-----+-----+-----+-----+-----+ 120
CAAGGCGCACCTAGATACAGGAGAACCCAACTGTCTTAACTCCTTCGTCTTTATTAGTA
10V P R G S M S S W V D R I E E A E I N H 29
↑ Thrombin cutting site

GAAAAATTTCACTCAGATCCACAGTACTTAAGGACTGAATTTAATTTACCTAAGATGGTA
121 -----+-----+-----+-----+-----+-----+ 180
CTTTTTAAAGTGAGTCTAGGTGTCATGAATTCCTGACTTAAATTAATGGATTCTACCAT
30E K F H S D P Q Y L R T E F N L P K M V 49

GCAGAAGAGATAAGACGAAAATGCCAGTATGCAGAATCATAGGAGAACAACCTGGGAGGA
181 -----+-----+-----+-----+-----+-----+ 240
CGTCTTCTCTATTCTGCTTTTACGGGTCATACGTCTTAGTATCCTCTTGTTCCACCCTCCT
50A E E I R R K C P V C R I I G E Q V G G 69

CAATTGAAAATAGGGCCTGGTATCTGGCAAATGGATTGCACACACTTTGATGGCAAATA
241 -----+-----+-----+-----+-----+-----+ 300
GTTAACTTTTATCCCGGACCATAGACCGTTTACCTAACGTGTGTGAACTACCGTTTTAT
70Q L K I G P G I W Q M D C T H F D G K I 89

ATTCTTGTGGGTATACATGTGGAATCAGGATATATATGGGCACAAATAATTTCTCAAGAA
301 -----+-----+-----+-----+-----+-----+ 360
TAAGAACACCCATATGTACACCTTAGTCCTATATATACCCGTGTTTATTAAAGAGTTCTT
90I L V G I H V E S G Y I W A Q I I S Q E 109

ACTGCTGACTGTACAGTTAAAGCTGTTTTACAATTGTTGAGTGCTCATAATGTTACTGAA
361 -----+-----+-----+-----+-----+-----+ 420
TGACGACTGACATGTCAATTTGACAAAATGTTAACAACCTCACGAGTATTACAATGACTT
110T A D C T V K A V L Q L L S A H N V T E 129

TTACAAACAGATAATGGACCAAATTTTAAAAATCAAAGATGGAAGGAGTACTCAATTAC
421 -----+-----+-----+-----+-----+-----+ 480
AATGTTTGTCTATTACCTGGTTTTAAATTTTGTAGTTTTCTACCTTCCTCATGAGTTAATG
130L Q T D N G P N F K N Q K M E G V L N Y 149

Fig. 8A

SUBSTITUTE SHEET (RULE 26)

9/13

ATGGGTGTGAAACATAAGTTTGGTATCCCAGGGAACCCACAGTCACAAGCATTAGTTGAA
481-----+-----+-----+-----+-----+-----+-----+ 540
TACCCACACTTTGTATTCAAACCATAGGGTCCCTTGGGTGTCAGTGTTCGTAATCAACTT
150 M G V K H K F G I P G N P Q S Q A L V E 169

AATGTAAATCATACATTAAAAGTTTGGATTGAGAAATTTTGCCTGAAACAACCTCCTTG
541-----+-----+-----+-----+-----+-----+-----+ 600
TTACATTTACTATGTAATTTTCAAACCTAAGTCTTTAAAAACGGACTTTGTTGGAGGAAC
170 N V N H T L K V W I Q K F L P E T T S L 189

GATAATGCCTTATCTCTCGCTGTACATAGTCTCAATTTTAAAAGAAGAGGTAGGATAGGA
601-----+-----+-----+-----+-----+-----+-----+ 660
CTATTACGGAATAGAGAGCGACATGTATCAGAGTTAAATTTTCTTCTCCATCCTATCCT
190 D N A L S L A V H S L N P K R R G R I G 209

GGGATGGCCCCTTATGAATTATTAGCACACAAGAATCCTTAAGAATACAAGATTATTTT
661-----+-----+-----+-----+-----+-----+-----+ 720
CCCTACCCGGAATACTTAATAATCGTGTGTTCTTAGGAATTCTTATGTTCTAATAAAA
210 G M A P Y R L L A Q Q E S L R I Q D Y F 229

TCTGCAATACCACAAAAATTGCAAGCACAGTGGATTTATTATAAAGATCAAAAAGATAAG
721-----+-----+-----+-----+-----+-----+-----+ 780
AGACGTTATGGTGTGTTTTAACGTTCTGTGCACCTAAATAATTTCTAGTTTTTCTATTCT
230 S A I P Q K L Q A Q W I Y Y K D Q K D K 249

AAATCGAAAGGACCAATGAGAGTAGAATACTGGGGACAGGGATCAGTATTATTAAAGGAT
781-----+-----+-----+-----+-----+-----+-----+ 840
TTTACCTTTCCTGGTTACTCTCATCTTATGACCCCTGTCCCTAGTCATAATAATTTCTTA
250 K W K C P M R V E Y W G Q G S V L L K D 269

GAAGAGAAGGGATATTTTCTTATACCTAGGAGACACATAAGGAGAGTTCCAGAACCCTGC
841-----+-----+-----+-----+-----+-----+-----+ 900
CTTCTCTTCCCTATAAAAGAATATGGATCCTCTGTGTATTCTCTCAAGGTCTTGGGACG
270 E E K G Y F D I P R R H I R R V P E P C 289

GCTCTTCCTGAAGGGGATGAGGGTACCAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTT
901-----+-----+-----+-----+-----+-----+-----+ 960
CGAGAACCACCTTCCCCTACTCCCATGGTTTCGCAATTGCCGGTCCGTTGTTCTCCACAAA
290 A L P E C D E G T K A L T A R Q Q E V F 309

FIV-IN (1-281) ↑ Linker ↑ E. coli Lex A (2-202)

GATCTCATCCGTGATCACATCAGCCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCG
961-----+-----+-----+-----+-----+-----+-----+ 1020
CTAGAGTAGGCACTAGTGTAGTCGGTCTGTCCATACGGCGGCTGGGCACGCCCTTTAGCGC
310 D L I R D H I S Q T G M P P T R A E I A 329

Fig. 8B

SUBSTITUTE SHEET (RULE 26)

10/13

1021-----+-----+-----+-----+-----+-----+1080
GTCGCAAACCCCAAGGCAAGGGGTTTGCGCCGACTTCTTGTAGACTTCCGCGACCGTGCG
330 Q R L G F R S P N A A E E H L K A L A R 349

AA A T
AAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCGGGATTTCGTCTGTTGCAGGAAGAG
1081-----+-----+-----+-----+-----+-----+1140
TTTCCGCAATAACTTTAACAAAGGCCGCGTAGTGCGCCCTAAGCAGACAACGTCCTTCTC
350 K G V I E I V S G A S R G I R L L Q E E 369

GAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAG
1141-----+-----+-----+-----+-----+-----+1200
CTTCTTCCCAACGGCGACCATCCAGCACACCGACGGCCACTTGGTGAAGACCGCGTTGTC
370 E E G L P L V G R V A A G E P L L A Q Q 389

CATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAACCCGAATGCTGATTTCCTG
1201-----+-----+-----+-----+-----+-----+1260
GTATAACTTCCAGTAATAGTCCAGCTAGGAAGGAATAAGTTCCGGCTTACGACTAAAGGAC
390 H I E G H Y Q V D P S L F K P N A D F L 409

CTGCGCGTCAGCGGGATGTCGATGAAAGATATCGGCATTATGGATGGTGACTTGTGGCA
1261-----+-----+-----+-----+-----+-----+1320
GACGCGCAGTCGCCCTACAGCTACTTTCTATAGCCGTAATACCTACCACTGAACGACCGT
410 L R V S G M S M K D I G I M D G D L L A 429

GTGCATAAACTCAGGATGTACGTAACGGTCAGGTCGTTGTGCGACGTATTGATGACGAA
1321-----+-----+-----+-----+-----+-----+1400
CACGTATTTTGAAGTCCTACATGCATTGCCAGTCCAGCAACAGCGTGCATAACTACTGCTT
430 V H K T Q D V R N G Q V V V A R I D D E 449

GTTACCGTTAAGCGCCTGAAAAACAGGGCAATAAAGTCGAACCTGTTGCCAGAAAATAGC
1401-----+-----+-----+-----+-----+-----+1460
CAATGGCAATTTCGCGGACTTTTTTGTCCCGTTATTTTCAGCTTGACAACGGTCTTTTATCG
450 V T V K R L K K Q G N K V E L L P E N S 469

GAGTTTAAACCAATTGTCGTTGACCTTCGTCAGCAGAGCTTCACCATTGAAGGGCTGGCG
1461-----+-----+-----+-----+-----+-----+1520
CTCAAATTTGGTTAACAGCAACTGGAAGCAGTCGTCTCGAAGTGGTAACCTCCCGACCGC
470 E F K P I V V D L R Q Q S F T I E G L A 489

GTTCCGGTTATTTCGCAACGGCGACTGGCTGTAAGGATCC
1521-----+-----+-----+-----+-----+-----+1559
CAACCCCAATAAGCGTTGCCGCTGACCGACATTCTAGG
490 V G V I R N G D W L *** 499

Fig. 8C

SUBSTITUTE SHEET (RULE 26)

11/13

ATTTTGTTTAACTTTAAGAAGGAGATATACATAATGCATCACCATCACCATCACCATATC
1-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 60
TAAACAAATTGAAATTCTTCCTCTATATGTATTACGTAGTGGTAGTGGTAGTGGTATAG
M H H H H H H H I 9
His-Tag

GTTCCGCGTGGATCTATGTCCTCTTGGGTTGACAGAATTGAGGAAGCAGAAATAAATCAT
61-----+-----+-----+-----+-----+-----+-----+-----+-----+ 120
CAAGGCGCACCTAGATACAGGAGAACCAACTGTCTTAACTCCTTCGTCTTTATTTAGTA
10V P R G S M S S W V D R I E E A E I N H 29
↑ Thrombin cutting site

GAAAAATTTCACTCAGATCCACAGTACTTAAGGACTGAATTTAATTTACCTAAGATGGTA
121-----+-----+-----+-----+-----+-----+-----+-----+-----+ 180
CTTTTTAAAGTGAGTCTAGGTGTCATGAATTCCTGACTTAAATTAATGGATTCTACCAT
30E K F H S D P Q Y L R T E F N L P K M V 49

GCAGAAGAGATAAGACGAAAATGCCAGTATGCAGAATCATAGGAGAACAACTGGGAGGA
181-----+-----+-----+-----+-----+-----+-----+-----+-----+ 240
CGTCTTCTCTATTCTGCTTTTACGGGTCATACGTCTTAGTATCCTCTTGTTCCACCTCCT
50A E E I R R K C P V C R I I G E Q V G G 69

CAATTGAAAATAGGGCCTGGTATCTGGCAAATGGATTGCACACACTTTGATGGCAAAATA
241-----+-----+-----+-----+-----+-----+-----+-----+-----+ 300
GTAACTTTTATCCCGGACCATAGACCGTTTACCTAACGTGTGTGAACTACCGTTTTAT
70Q L K I G P G I W Q M D C T H F D G K I 89

ATTCTTGTGGGTATACATGTGGAATCAGGATATATATGGGCACAAATAATTTCTCAAGAA
301-----+-----+-----+-----+-----+-----+-----+-----+-----+ 360
TAAGAACACCCATATGTACACCTTAGTCCTATATATACCCGTGTTTATTAAAGAGTTCTT
90I L V G I H V E S G Y I W A Q I I S Q E 109

ACTGCTGACTGTACAGTTAAAGCTGTTTTACAATTGTTGAGTGCTCATAATGTTACTGAA
361-----+-----+-----+-----+-----+-----+-----+-----+-----+ 420
TGACGACTGACATGTCAATTTGACAAAATGTTAACAACCTCAGGAGTATTACAATGACTT
110T A D C T V K A V L Q L L S A H N V T E 129

TTACAAACAGATAATGGACCAAATTTTAAAAATCAAAGATGGAAGGAGTACTCAATTAC
421-----+-----+-----+-----+-----+-----+-----+-----+-----+ 480
AATGTTTGTCTATTACCTGGTTTTAAATTTTGTAGTTTTCTACCTTCCTCATGAGTTAATG
130L Q T D N G P N F K N O K M E G V L N Y 149

Fig. 9A

SUBSTITUTE SHEET (RULE 26)

12/13

ATGGGTGTGAAACATAAGTTTGGTATCCCAGGGAACCCACAGTCACAAGCATTAGTTGAA
481-----+-----+-----+-----+-----+ 540
TACCCACACTTTGTATTCAAACCATAGGGTCCCTTGGGTGTCAGTGTTTCGTAATCAACTT
150 M G V K H K F G I P G N P Q S Q A L V E 169

AATGTAATCATACATTAAAAGTTTGGATTGAGAAATTTTTGCCTGAAACAACCTCCTTG
541-----+-----+-----+-----+-----+ 600
TTACATTTAGTATGTAATTTTCAAACCTAAGTCTTTAAAAACGGACTTTGTTGGAGGAAC
170 N V N H T L K V W I Q K F L P E T T S L 189

GATAATGCCTTATCTCTCGCTGTACATAGTCTCAATTTTAAAAGAAGAGGTAGGATAGGA
601-----+-----+-----+-----+-----+ 660
CTATTACGGAATAGAGAGCGACATGTATCAGAGTTAAATTTTCTTCTCCATCCTATCCT
190 D N A L S L A V H S L N F K R R G R I G 209

GGGATGGCCCCTTATGAATTATTAGCACAAACAAGAATCCTTAAGAATACAAGATTATTTT
661-----+-----+-----+-----+-----+ 720
CCCTACCGGGGAATACTTAATAATCGTGTGTTGTTCTTAGGAATTCTTATGTTCTAATAAAA
210 G M A P Y E L L A Q Q E S L R I Q D Y F 229

TCTGCAATACCACAAAAATTGCAAGCACAGTGGATTTATTATAAAGATCAAAAAGATAAG
721-----+-----+-----+-----+-----+ 780
AGACGTTATGGTGTGTTTTAACGTTTCGTGTACCTAAATAATATTTCTAGTTTTTCTATTC
230 S A I P Q K L Q A Q W I Y Y K D Q K D K 249

AAAGGTACCAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTT GATCTCATCCGTGATCAC
781-----+-----+-----+-----+-----+ 840
TTTCCATGGTTTCGCAATTGCCGGTCCGTTGTTCTCCACAACTAGAGTAGGCACTAGTG
250 K G T K A L T A R Q Q E V F D L I R D H 269
↑ Linker ↑ E. coli Lex A (2-202)
FIV-IN (1-235)

ATCAGCCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAGCGTTTGCGGTTCCGT
841-----+-----+-----+-----+-----+ 900
TAGTCGGTCTGTCCATACGGCGGCTGCGCACGCCTTTAGCGCGTCGCAACCCCAAGGCA
270 I S Q T G M P P T R A E I A Q R L G F R 289

TCCCCAAACGCGGCTGAAGAACATCTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATT
901-----+-----+-----+-----+-----+ 960
AGGGCTTTGCGCCGACTTCTTGTAAGTCTCCGCAACCGTGCGTTTCCGCAATAACTTTAA
290 S F N A A E E H L K A L A R K G V T E I 309

Fig. 9B
SUBSTITUTE SHEET (RULE 26)

13/13

961 GTTCCGGCGCATCACGCGGGATTCTGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTG
-----+-----+-----+-----+-----+1020
CAAAGGCCCGTAGTGCGCCCTAAGCAGACAACGTCCTTCTCCTTCTTCCCAACGGCGAC
310 V S G A S R G I R L L Q E E E E G L P L 329

1021 GTAGGTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAGCATATTGAAGGTCATTAT
-----+-----+-----+-----+-----+1080
CATCCAGCACACCGACGGCCACTTGGTGAAGACCGCGTTGTCGTATAACTTCCAGTAATA
330 V G R V A A G E P L L A Q Q H I E G H Y 349

1081 CAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTCTGCTGCGCGTCAGCGGGATG
-----+-----+-----+-----+-----+1140
GTCCAGCTAGGAAGGAATAAGTTGGGCTTACGACTAAAGGACGACGCGCACTCGCCCTAC
350 Q V D P S L F K P N A D F L L R V S G M 369

1141 TCGATGAAAGATATCGGCATTATGGATGGTGACTTGCTGGCAGTGCATAAACTCAGGAT
-----+-----+-----+-----+-----+1200
AGCTACTTTCTATAGCCGTAATACCTACCACTGAACGACCGTCACGTATTTTGAGTCCTA
370 S M K D I G I M D G D L L A V H K T Q D 389

1201 GTACGTAACGGTCAGGTCGTTGTGCGCACGTATTGATGACGAAGTTACCGTTAAGCGCCTG
-----+-----+-----+-----+-----+1260
CATGCATTGCCAGTCTAGCAACAGCGTGCATAACTACTGCTTCAATGGCAATTCGCGGAC
390 V R N G Q V V V A R I D D E V T V K R L 409

1261 AAAAAACAGGGCAATAAAGTCGAACTGTTGCCAGAAAATAGCGAGTTTAAACCAATTGTC
-----+-----+-----+-----+-----+1320
TTTTTTGTCCCGTTATTTTCAGCTTGACAACGGTCTTTTATCGCTCAAATTTGGTTAACAG
410 K K Q G N K V E L L P E N S E F K P I V 429

1321 GTTGACCTTCGTCAGCAGAGCTTCACCAATTGAAGGGCTGGCGGTTGGGGTTATTCGCAAC
-----+-----+-----+-----+-----+1380
CAACTGGAAGCAGTCGTCTCGAAGTGGAACCTTCCCGACCGCCAACCCCAATAAGCGTTG
430 V D L R Q Q S F T I E G L A V G V I R N 449

1381 GGCGACTGGCTGTAAGGATCC
-----+-----+1401
CCGCTGACCGACATTCCTAGG
450 G D W L *** 453

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/19277

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/00; C07H 19/00, 21/00; C07K 1/00, 14/00, 17/00

US CL : 435/183; 530/350; 536/22.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/183; 530/350; 536/22.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P ----- Y, P	WO 96/35777 A1 (FOX CHASE CANCER CENTER) 14 November 1996, page 1, abstract.	1-8, 10, 12, 13, 15-19, 22-30 ----- 9, 11, 14, 20, 21, 31
X, P ----- Y, P	GOULAOUIC et al. Directed Integration of viral DNA Mediated by Fusion Protein Consisting of Human Immunodeficiency Virus Type 1 Integrase and Escherichia coli LexA Protein. Journal of Virology, January 1996, Vol. 70, No. 1, pages 37-46, see abstract.	1-4, 6-8, 10, 12, 13, 15-19, 22, 24, 30 ----- 5, 9, 11, 14, 20, 21, 23, 31

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 24 FEBRUARY 1997	Date of mailing of the international search report 27 MAR 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer <i>I W for</i> NASHAAT T. NASHED Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

 International application No.
 PCT/US96/19277

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, E ----- Y, E	BUSHMAN et al. Tethering Human Immunodeficiency Virus Type 1 Preintegration Complexes to Target DNA Promotes Integration at Nearby Sites. Journal of Virology, January 1997, Vol. 71, pages 458-464, see abstract.	1-4, 10, 12, 13, 15-19, 22, 24, 29 ----- 5-9, 14, 20, 21, 23, 25-28, 30, 31
X, P ----- Y, P	KATZ et al. Targeting of Retroviral Integrase by Fusion to a Heterologous DNA Binding Domain: In Vitro Activities and Incorporation of a Fusion Protein into Viral Particles. Virology. 01 March 1996, Vol. 217, No. 1, pages 178-190, see abstract.	1, 6-8, 10, 12, 13, 15-19, 22, 25-28 ----- 2-5, 9, 11, 14, 20, 21, 23, 24, 29-31.
X --- Y	BUSHMAN. Targeting Retroviral Integration. Science. 10 March 1995, Vol. 267, pages 1443-1444, the entire document.	1-4, 10, 12, 13, 15-19, 22, 24 ----- 5, 6-9, 11, 14, 20, 21, 23, 25-31
X --- Y	BUSHMAN. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA. September 1994, Vol. 91, pages 9233-9237, see abstract.	1-4, 10, 12, 13, 15-19, 22, 24 ----- 5-9, 11, 14, 20, 21, 23, 25-31
Y	VINK et al. Activities of the Feline Immunodeficiency Virus Integrase Protein Produced in Escherichia coli. Journal of Virology, March 1994, Vol. 68, No. 3, pages 1468-1474, see abstract.	1, 5, 12, 13, 15- 19, 22, 24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/19277

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/19277

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN: Medline, Caplus, Biosis, Embase, WPIDS, and Biotechchds.

Search terms: Integrase, fusion protein, antisense, oligonucleotide, protein-oligonucleotide, and catalytic domain.